Do You Feel the Same? The Effect of Outcome Severity on Moral Judgment and Interpersonal Goals of Perpetrators, Victims, and Bystanders

Lisa K. Frisch[1], Markus Kneer[1], Joachim I. Krueger[2], Johannes Ullrich[1]

[1] University of Zürich

[2] Brown University

Version: 17.01.2020

Author Note: Correspondence concerning this article should be addressed to Lisa Frisch, University of Zurich, Binzmühlestrasse 14/15, 8057 Zürich, Switzerland. E-mail: *lisakatharina.frisch@uzh.ch*

Abstract

When two actors have exactly the same mental states but one happens to harm another person (unlucky actor) and the other one does not (lucky actor), the latter elicits milder moral judgment among bystanders. We hypothesized that the social role from which transgressions are perceived would moderate this outcome effect. In three preregistered experiments (N = 950), we randomly assigned participants to imagine and respond to moral scenarios as actor (i.e., perpetrator), victim, or bystander. Results revealed highly similar outcome effects on moral judgment across social roles. However, as predicted, the social role moderated the strength of the outcome effect on interpersonal goals pertaining to agency and communion. Although in agreement about the blameworthiness of lucky and unlucky actors, victims' agency and communion were more sensitive to the outcome severity than perpetrators' agency and communion, with bystanders' outcome sensitivity falling in between. Outcome severity affected agency and communion directly instead of being mediated by moral judgment. We discuss the possibility that outcome severity raises normative expectations regarding interaction in a transgression's aftermath that are unrelated to moral considerations.

Keywords: outcome effect, moral judgment, moral luck, social role, needs-based model, agency, communion

Do You Feel the Same? The Effect of Outcome Severity on Moral Judgment and Interpersonal Goals of Perpetrators, Victims, and Bystanders

How laypeople make moral judgments is a question that concerns not only moral philosophers but also anyone who is familiar with the feeling that they have been wronged. Philosophers wonder if their ethical maxims resonate with people's intuitions. Everyone else also hopes that their own moral judgment is shared by other people. After all, it would make for a smoother resolution of conflicts if victims and perpetrators could agree on the amount of blame an action deserves, and if bystanders were to further validate their judgments. Considering that moral judgment seems likely to influence how the parties will interact with each other in the aftermath of a transgression, the consensus of moral judgment appears all the more desirable. In the present research, we focus on one particular determinant of moral judgment: the severity of a transgression's outcome. Curiously enough, there exists no research that would have compared outcome effects across victims, perpetrators and bystanders or examined whether their moral judgment are related to the interpersonal goals of these different parties. We present the findings of three preregistered studies testing the proposition that victims are the most sensitive to outcome information and perpetrators the least.

Does Outcome Severity Affect Moral Judgment – and Should It?

Imagine, as Adam Smith did in his Theory of Moral Sentiments (1759), a man throwing a stone over a tall wall onto a public street, killing a person. This man, we would be inclined to judge, deserves to be punished. At the same time, Smith observed that nothing "would appear more shocking to our natural sense of equity, than to bring a man to the scaffold merely for

having thrown a stone carelessly into the street without hurting anybody. The folly and inhumanity of his conduct, however, would in this case be the same; but still our sentiments would be every different" (p. 152).

Empirical research shows that the severity of an outcome affects moral judgment. People judge an action as less permissible, more wrong, more blameworthy, and call for greater punishment when the action leads to greater harm, even if this harm was beyond the agent's control (Cushman, 2008; Cushman, Dreber, Wang, & Costa, 2009; Gino, Moore, & Bazerman, 2009; Gino, Shu, & Bazerman, 2010; Kneer & Machery, 2019; Lench, Domsky, Smallman, & Darbor, 2015; Young, Nichols, & Saxe, 2010; for a review, cf. Martin & Cushman, 2015).

We can ask whether outcome severity *should* affect moral judgment. Two positions mark the endpoints of a spectrum in normative ethics: The first position is *Deontology*, according to which "nothing in the world […] could be called good without qualification except a good will" (Kant, 1785, p.1). According to this view, an action taken for dubious reasons is wrong despite a beneficial outcome. The second position is *Consequentialism,* which regards an action as good to the extent that its consequences increase well-being. Thus, regardless of the agent's intentions, radical consequentialists would judge one and the same action as morally better, the better the outcome. Both positions, and individuals' willingness to toggle between them are familiar from work on the trolley problem (Foot 1967; Thomson 1976, 1985). Whereas with deontology morality depends exclusively on the agent's intentions (or, broadly conceived, their mental states), with consequentialism only the outcomes matter.

Whereas the outcome effect is consistent with consequentialist ethics, it clashes with a fundamental maxim of Kantian ethics, the Control Principle, according to which agents are only

responsible for features pertaining to a sequence of action which are under their control. The Control Principle suggests that throwing the stone deserves a fixed amount of blame, whether the stone happens to fatally hit a person or not. Throwing the stone even deserves blame if it, by some unexpected turn of events, produces a beneficial, yet not envisioned, outcome. At the same time, many people are inclined towards a marked difference in moral evaluation across the two cases discussed by Smith. This intuition, running counter to the Control Principle, generates the *Puzzle of Moral Luck*, which has been at the heart of extensive debate in moral philosophy (Nagel, 1979; Williams, 1981; for recent work on Moral Luck, see also Nelkin, 2019; Hartman, 2017).

The difference between Kantian and consequentialist ethics is reflected in individual differences in outcome sensitivity. In fact, the outcome effect is sometimes exhibited only by a small proportion of participants while most people fail to see a moral difference between actions ending in good or bad outcomes (e.g., Schwitzgebel & Cushman, 2012; Kneer & Machery, 2019). Such heterogeneity of effects calls for an examination of potential moderator variables. In the present research, we focus on the social role of the judge. While previous research has exclusively used scenarios in which uninvolved bystanders judge incidents, it is possible that judges spontaneously identify with the victim or the perpetrator. Potential moderating effects of social role are also practically relevant. From a restorative justice perspective (Wenzel, Okimoto, Feather, & Platow, 2008), it is important that perpetrators and victims go beyond retribution and engage in interactions that symbolically undo the transgression. Such dialogue and interaction between perpetrators and victims would be greatly informed by an understanding of the role-specific moral judgments.

Since our interests in moral judgment (on which the literature is plentiful) ultimately regard interactions in the transgression's aftermath, we also extend the research on outcome effects from moral judgments to moral action. Specifically, we (i) report the first experiments that take the social role (perpetrator vs. victim vs. bystander) of the moral assessor into account, and (ii) explore how it impacts the assessor's goals for interacting with the other parties involved.

In the following, we first review the explanations that the psychological literature offers to understand outcome effects among neutral bystanders. Outcome effects across different hypothetical social roles are explored with reference to introspection illusion (Pronin, 2009), which suggest that victims will be more prone to the outcome effect than perpetrators. Drawing on the needs-based model of reconciliation (Shnabel & Nadler, 2015), we further hypothesize that the difference in judgments across the perpetrator and victim roles extend to their interpersonal goals. Human interaction can be parsimoniously described in terms of the broad dimensions of agency and communion (Abele & Wojciszke, 2014; Bakan, 1966; Fiske, Cuddy, Glick, & Xu, 2002; Locke, 2015; Reeder & Brewer, 1979; Wiggins, 1979, 1991). Agentic goals include seeking control, dominance, and power, whereas communal goals include caring for, cooperating with, and being connected with others. Given that victims often pursue agentic goals and perpetrators often pursue communal goals (Shnabel & Nadler, 2015), outcome severity might also influence the role-specific goals for agency and communion.

Finally, a clarification is in order regarding our use of the term "perpetrator" in the context of moral luck. Although philosophers or legal experts are likely to disagree as to whether the term "perpetrator" is appropriate for an actor who is putting others at risk to be harmed, previous research on the social role of perpetrators may inform our understanding of the outcome effect.

Laypeople judging such incidents may or may not make the same distinctions as philosophers or legal experts. Thus, we will refer to both lucky and unlucky actors as perpetrators in order to be able to merge the different literatures.

Explanations of the Outcome Effect

The most straightforward model hypothesizes a direct effect of outcome on moral judgment (the Simple View). According to a second type of model, different outcome types trigger different inferences regarding the epistemic states (knowledge and belief) of the agent, which mediate the effect of outcome on moral judgment. Severe outcomes lead the judging subject to infer that the agent must have known his action would produce harmful consequence (Royzman & Kumar, 2004; Young, Scholz, & Saxe, 2011; for discussion see Nichols, Timmons, Lopez, 2014; and Kamtekar & Nichols, 2019). The impact of outcome on moral judgment would thus be mediated by epistemic state ascriptions. Another mediation account suggests that severity of outcome affects the perceived likelihood of the harmful outcome, which leads to differing ascriptions of *mens rea* (or "inculpating state of mind") defined in terms of risk such as recklessness or negligence, and consequently moral judgment (Kneer & Machery, 2019).

These hypothesized mechanisms all allow for a rationalist interpretation or else can be viewed as a manifestation of biased judgment. One might, for instance, consider a harmful outcome that just occurred good evidence for its likelihood. A higher likelihood warrants the ascription of negligence (i.e., that the agent should have been aware of a substantial risk, cf. Amaya, forthcoming, and Model Penal Code 2.02), and hence – given the presence of a *mens rea* – the ascription of more blame in the bad outcome case than in the neutral one (rationalist epistemic state mediation view; Thomson, 1993; Rosebury, 1995; Young et al., 2010; Alfano et

al., 2014). Alternatively, one could interpret the different assessments of probability across outcomes as an instance of the hindsight bias (Fischhoff, 1975): *ex-post* considerations inappropriately interfere with evaluations that should be undertaken from an *ex-ante* perspective. The biased assessment of the outcome probability leads to a distorted evaluation of *mens rea* and consequently to an inappropriate moral judgment in the bad outcome case. (Royzman & Kumar, 2004; Alicke, 2000; Alicke & Rose, 2010; Nadelhoffer, 2006; Kneer & Machery, 2019).

Outcome Effects Across Social Roles

How will social role impact the outcome effect? According to Schein and Gray (2018), harm is a matter of perception, and perceived harm can differ from the vantage points of perpetrators, victims, and bystanders. For one thing, the orthogonal distinction between first-person and third-person perspective underpins an asymmetry in epistemic access. In the actor perspective, we have direct access to our mental states and motives. In the observer perspective, we infer others' intentions, beliefs and other mental states from behavioral cues. The epistemic authority we enjoy from the first-person perspective is frequently deceptive (on misperceptions regarding one's own intentions, cf. Epley & Dunning, 2000; regarding emotions, cf. Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998; concerning prescient thoughts, cf. Pronin, Wegner, McCarthy, & Rodriguez, 2006). Importantly, however, the asymmetry in epistemic access engenders an introspection illusion (Pronin, 2009) with two components: In self-assessment, people tend towards disproportionally heavy weighting of introspectively available information (i.e. mental state information and personal feelings). In other-assessment, they gravitate towards a disproportionally heavy weighting of extrospectively available evidence (i.e. behavioral and outcome information; see also Malle, 2005). Notably, this perspective-specific weighting of

evidence persists even when actors and observers are equally provided with both introspective and extrospective information about behavior (Pronin & Kugler, 2007).

The introspection illusion suggests the following hypotheses regarding the outcome effect across social roles: Unlike perpetrators, who not only observe the outcome but also have access to their mental states prior to the outcome, victims and bystanders, in their assessment of the actor from an observer perspective, can be expected to put excessive weight on the outcome, while attributing little weight to the perpetrator's mental states. For both bystanders and victims, we can expect the outcome effect to be stronger than in the perpetrators' self-assessment. For victims, there is some evidence of this pattern: They tend to exaggerate the negative consequences of unfair treatment and view them as a persistent pattern of misbehavior on the perpetrator's behalf (Baumeister, 1999; Baumeister, Stillwell, & Wotman, 1990). But we can also expect a difference between victims and bystanders: Since victims are involved in the focal action (Adams & Inesi, 2017; Jones & Nisbett, 1972) and directly affected by the action's outcome (McCullough, Fincham, & Tsang, 2003), their tendency to base their judgment on the severity of outcome should be stronger than that of bystanders. In short, we predict the outcome effect to be larger among victims than among bystanders, and to be weaker among perpetrators than among bystanders.

Agentic and Communal Interpersonal Goals

We draw on the needs-based model (Shnabel & Nadler, 2008; Shnabel & Nadler, 2015) to understand how dispositions for moral action emerge. According to this model, moral transgressions asymmetrically threaten victims' and perpetrators' identities and elicit different needs. To the extent that they view an action as immoral, the needs-based model assumes

perpetrators to suffer from a threat to their identity as moral actors, whereas victims are assumed to suffer from a threat to their identity as agentic actors. As most people seek to maintain positive identities, both victims and perpetrators experience diverging needs. Perpetrators should feel a heightened need for social acceptance, whereas victims should experience a heightened need for empowerment. To restore their threatened identities, perpetrators pursue communal goals (communion refers to "the self as part of the community and is geared toward closeness, affection, and cooperation"; Grosse Holtforth, Thomas, & Caspar, 2011, p. 109), whereas victims pursue agentic goals (agency refers to "the pursuit of independence and autonomy of the individual and aims at control, assertiveness and self-enhancement"; Grosse Holtforth et al., 2011, p. 109).

An important moderator of the effects of social role on agentic and communal goals is the extent to which the conflict parties view the action as immoral. For example, if perpetrators deny their perpetrator role, the model does not predict increased communion. In fact, in many protracted conflicts there is competition for the victim role (e.g., Noor, Shnabel, Halabi, & Nadler, 2012) and where there is ambiguity with regard to the roles of victim and perpetrator, the agentic goals associated with the victim role are more pronounced (SimanTov-Nachlieli, Shnabel, Aydin, & Ullrich, 2018). Thus, the assumption bridging the need-based model with the moral luck literature is that the harshness of moral judgment, capturing perpetrator's moral debt to the victim, should determine whether and to what extent the perpetrator has communal goals vis-à-vis the victim and the victim has agentic goals vis-à-vis the perpetrator. If so, outcome severity, which we assume to affect moral judgment, should have implications for the interpersonal goals of perpetrators and victims as well. More specifically, victims' interpersonal goals should also

vary more strongly between situations with negative and neutral outcomes than perpetrators' interpersonal goals. The predominant goal of victims would be to act agentically toward the perpetrator, whereas the predominant goal of the perpetrator would be to act communally toward the victim. We propose that bystanders share the perspective of victims when interacting with the perpetrator, thus responding with victim-type interpersonal goals (i.e., agency) in proportion to their moral judgments. In contrast, when bystanders interact with victims, we assume them to approach victims with communion, thus responding with perpetrator-type interpersonal goals.

Although not specified in the needs-based model, there is evidence that an increase in victim goals (i.e., agency) is accompanied by a decrease of perpetrator goals (i.e., communion) and vice versa. Research extending the needs-based model to the context of social class found that when people imagined interactions with members of a higher social class, they wanted to act more agentically *and* less communally compared with social interactions within their own social class. Conversely, when people imagined interactions with members of a lower social class, they wanted to act more communally *and* less agentically (Aydin, Ullrich, Siem, Locke, & Shnabel, 2019). It is plausible that such patterns are also present in perpetrators' and victims' interactions after transgressions. Compared with bystanders, victims might simultaneously pursue more agency (i.e., their role-specific goal) and less communion, whereas perpetrators might simultaneously pursue more communion (i.e., their role-specific goal) and less agency.

The Present Research

We conducted three preregistered online studies to examine how social role moderates the outcome effect on moral judgment and resulting interpersonal goals (for preregistrations, data,

analyses, materials, and online appendix, see *https://osf.io/t7e2k/*). Study 1 contrasted participants' reactions to a moral scenario in the social roles of perpetrator and victim. Studies 2 and 3 were conducted to replicate the effects with a different moral scenario and to examine the prediction that the outcome effects among bystanders are weaker than among victims, but stronger than among perpetrators.

Our suggested explanation for the moderating effects of social role on the outcome effect assumes that the subjective, role-specific patterns arising in actual transgressions extend to vignette-based evaluations from imagined role-specific perspectives. Feltz, Harris, and Perez (2012) call this premise into question on the basis of several studies on epistemic state ascriptions, where actor-observer differences in an actual economic game do not replicate in imagined-role vignette studies. Kneer (2018) partially reinstates it and emphasizes that triggering role-specific imagination can require extensive preliminary exercises. However, there is some evidence from imagination-based experiments that suggests that the proposed hypotheses are on track: Stillwell and Baumeister (1997), for example, found that participants imagining themselves in a perpetrator role focused more heavily on mitigating circumstances and frequently invoked mental state information, whereas victims emphasized the severity of the offense. Studies by Exline, Yali, and Lobel (1998) as well as Adams and Inesi (2016; experiments 3 and 5) suggest that when individuals are placed into an imagined perpetrator instead of the victim role, they see their transgressions as less harmful, frequent, intentional, and malicious. Finally, Leunissen, De Cremer, Reinders Folmer, and Van Dijke (2013) found the same differences between perpetrators and victims regarding their needs for apologies across studies using remembered actual transgressions or scenarios that were only imagined. Thus, viewed in the context of similar

previous studies, the scenario-based methodology chosen for the present research appears to be appropriate.

Study 1

The objective of Study 1 was to investigate the outcome effect on moral judgment and interpersonal goals of the parties involved in a transgression (i.e. perpetrator and victim). Bad outcomes should elicit harsher moral judgments and stronger role-specific goals (i.e., agency for victims and communion for perpetrators) than neutral outcomes. Drawing on the introspection illusion (Pronin, 2009), we hypothesized that the social role in which transgressions are perceived would moderate the outcome effect. Given that outcome-related information will be factored in more strongly by victims than by perpetrators, victims should exhibit a stronger outcome effect than perpetrators. Note that we preregistered only effects on agency. To expand the preregistered pattern to moral judgments and communion seemed meaningful *post hoc*. However, the effects on the latter variables need to be interpreted with caution.

Method

To test our hypothesis, we conducted an online study, using vignettes for experimental manipulation. The study had a 2x2x2 design with outcome (neutral vs. bad) as a within-subjects factor and social role (perpetrator vs. victim) and mental states (unintended vs. intended) as between-subjects factors. To address outcome sensitivity among perpetrators and victims, we manipulated outcome severity within-subjects. This increases the manipulations' validity as it creates a stable comparative context rather than leaving it up to participants to construe a comparison for judging the experimental stimuli (Birnbaum, 1999). Moreover, manipulating the

factor of interest within-subjects decreases error variance, thereby increasing the statistical power (Birnbaum & Mellers, 1983). Although the focus of the study lies on the outcome effect, we additionally manipulated mental states, as prior research has shown that people tend to draw inferences from outcome severity to the agent's intentions (e.g., Fincham, 1982; for a discussion see Malle, 2014). By explicitly stating whether the actor intended to harm the victim we sought to rule out the possibility that participants use outcome severity as a proxy for intentions. Note, however, that mental states are merely a control variable and not part of our preregistered hypotheses.

Participants We recruited 116 participants by snowball sampling mainly among students at a large Swiss university.[1] Following our preregistered plan, we excluded all participants who (i) failed to correctly answer the comprehension check, (ii) failed to correctly answer the manipulation check, or (iii) did not complete the whole survey (for details see online appendix). The final sample included 107 participants (85 women, 21 men, one not specifying gender, Mdn$_{age}$ = 21, range 18 – 36). This sample size affords a power of .73 to detect a medium-sized effect (η_p^2= 0.06), assuming an alpha of 5%.

Procedure After providing informed consent, participants were randomly assigned to the between-subjects conditions of the factorial design. Participants then read two scenarios

[1] Additionally, we recruited 115 participants for an experimental design in which outcome severity was a between-subjects factor and mental states a within-subjects factor. However, as we did not preregister any hypothesis about the outcome effect in that design and the results do not inform the present manuscript, we do not report the results in detail.

(adopted from Cushman, 2008) with an action leading to a neutral or bad outcome. We held the order of the vignettes fixed, always beginning with the neutral outcome. Vignettes differed solely in their crucial words to manipulate treatment conditions. The following vignette was presented to participants in the perpetrator (victim) condition, when actions ended in a neutral outcome:

Imagine you are taking a class in sculpture. You are assigned to work with a partner to weld together pieces of metal. You do (Your partner does) not want to burn your partner's (your) hand. You (Your partner) only want(s) to weld together the metal. You do (Your partner does) not think that if you (he) weld(s) a piece of metal that your partner is (you are) holding, the heat will travel down the metal and burn your partner's (your) hand. You (Your partner) think(s) that the metal will weld without causing your partner (you) any injury at all. You (Your partner) weld(s) the metal but your partner happens (you happen) to let go and is (and are) not burned at all.

For vignettes that ended with a negative outcome solely the last sentence changed: *You (Your partner) weld(s) the metal and your partner's (your) hand is burned).*

Measures

Moral Judgment After each experimental vignette, participants rated the permissibility and wrongness of the action, and how much blame and punishment the actor deserved on scales from 1 to 5. We averaged the items wrongness, blame and (reverse-scored) permissibility after finding satisfactory internal consistency (Cronbach's $alpha = .81$). We analyzed punishment ratings separately because recent research suggests a unique effect of outcome on punishment, in that people are even more sensitive to outcome severity when judging

deserved punishment, compared to permissibility, wrongness, and blame (Kneer & Machery, 2019).

Interpersonal Goals To measure interpersonal goals, we used the German version of the Circumplex Scales of Intergroup Goals (CSIG; Locke, 2014), adapted to the interpersonal level as in Aydin et al. (2019). The CSIG consists of eight four-item scales each representing one octant in the circumplex (see Figure 1). Upper and lower octants along the vertical axis define high and low agency scores, whereas right and left octants along the horizontal axis define high and low communal scores. Each point within the circumplex represents a weighted mixture of agency and communion.

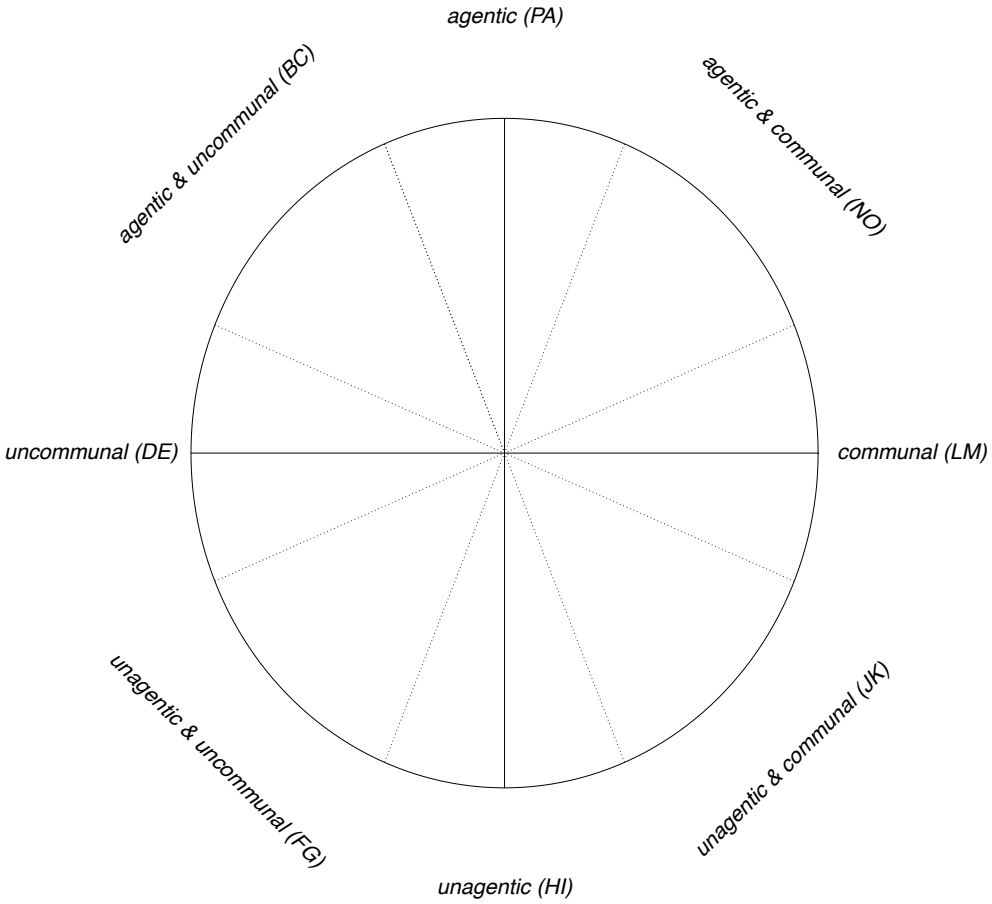

Figure 1. Circumplex of interpersonal goals (octants depicted with dotted lines).

After each experimental vignette (neutral or bad outcome), participants rated the 32 items of the CSIG. Specifically, they were asked to imagine an interaction with their partner from the experimental vignette and rate the respective goals from 1 = not important to 5 = very important: "When I interact with my partner it is important to me *that* … [e.g. I show concern for their welfare, I understand their point of view (example communal items) or I am assertive, I appear confident (example agentic items)". We calculated an overall agency score by combining the octant scores as follows: agentic goals = .414 × (PA – HI + (.707 × [BC + NO – JK – FG]; note

that each octant has by convention a generic two letter code). This score represents a participant's position on the vertical dimension in the circumplex shown in Figure 1. Likewise, we calculated an overall communal score by combining the octant scores as follows: communal goals = .414 × (LM – DE + (.707 × [JK + NO – BC – FG]; Leary, 1957; Locke 2014). This score represents a participant's position on the horizontal dimension in the circumplex shown in Figure 1. Cronbach's α were .85 for agentic goals and .95 for communal goals (see supplemental material for Cronbach's α of the octant scores).

Results[2]

Hypothesis Tests[3] We conducted 2(social role[perpetrator, victim]) x 2(outcome[neutral, bad]) x 2(mental state[unintended, intended]) ANOVAs with repeated measures on the second factor on moral judgment, punishment, agency and communion. Our key hypothesis addressed the moderating influence of social role on the outcome effect (i.e., an interaction between social role and outcome). We expected that victims would exhibit a stronger outcome effect than perpetrators.

[2] We used R (Version 3.4.3; R Core Team, 2017b) and the R-packages *afex* (Version 0.20.2; Singmann, Bolker, Westfall, & Aust, 2016), *psy* (Version 1.1; Falissard, 2012), emmeans (Version 1.4.2; Lenth, 2019), dplyr (Version 0.8.3; Wickham, François, Henry, & Müller, 2019), tidyr (Version 1.0.0; Wickham & Henry, 2019), and taRifx (Version 1.0.6.1; Friedman, 2018) for analyses of the three studies. We created the Figures using the R-package *ggplot2* (Wickham, 2016).

[3] Following our preregistration, we tested our hypotheses with alpha = 5%, one-sided tests, without adjustment for multiple comparisons.

Starting with moral judgment, there was a significant interaction effect of social role and outcome ($F(1,103) = 4.92$, $p = .029$, $\eta_p^2 = .05$, $d = .41$), indicating that, as predicted, victims showed stronger outcome effects than perpetrators. Table 1 shows means and standard deviations as well as test statistics associated with the simple main effects of outcome within each social role condition. As can be seen in the right column of Table 1, victims judged transgressions more harshly when resulting in a bad outcome than in a neutral outcome. By contrast, perpetrators' outcome effects on moral judgment were non-significant.

These diverging outcome effects are shown as the distributions of difference scores (moral judgment following bad outcome minus moral judgment following neutral outcome) in upper left-hand side of Figure 2. The expansion of the density plot represents the frequency of observations. Means and 95% confidence intervals are shown within the violin-shaped plots. Inspection of the plot further shows that in both social role conditions, the violin's expansion is greatest at the value of zero, which means that a typical response pattern (27 % of participants) was to make identical moral judgments across outcome conditions.

For punishment judgments, there was only a main effect of outcome, ($F(1,103) = 25.61$, $p < .001$, $\eta_p^2 = .2$). Participants reported actions that produced negative outcomes to deserve more punishment ($M = 2.96, SD = 1.28$) than actions that produced neutral outcomes ($M = 2.57, SD = 1.15$). The outcome x social role interaction was non-significant ($F(1,103) = 0.39$, $p = .53$, $\eta_p^2 = .00$). As illustrated in the upper right-hand side of Figure 2, perpetrators and victims exhibited similar outcome effects on punishment.

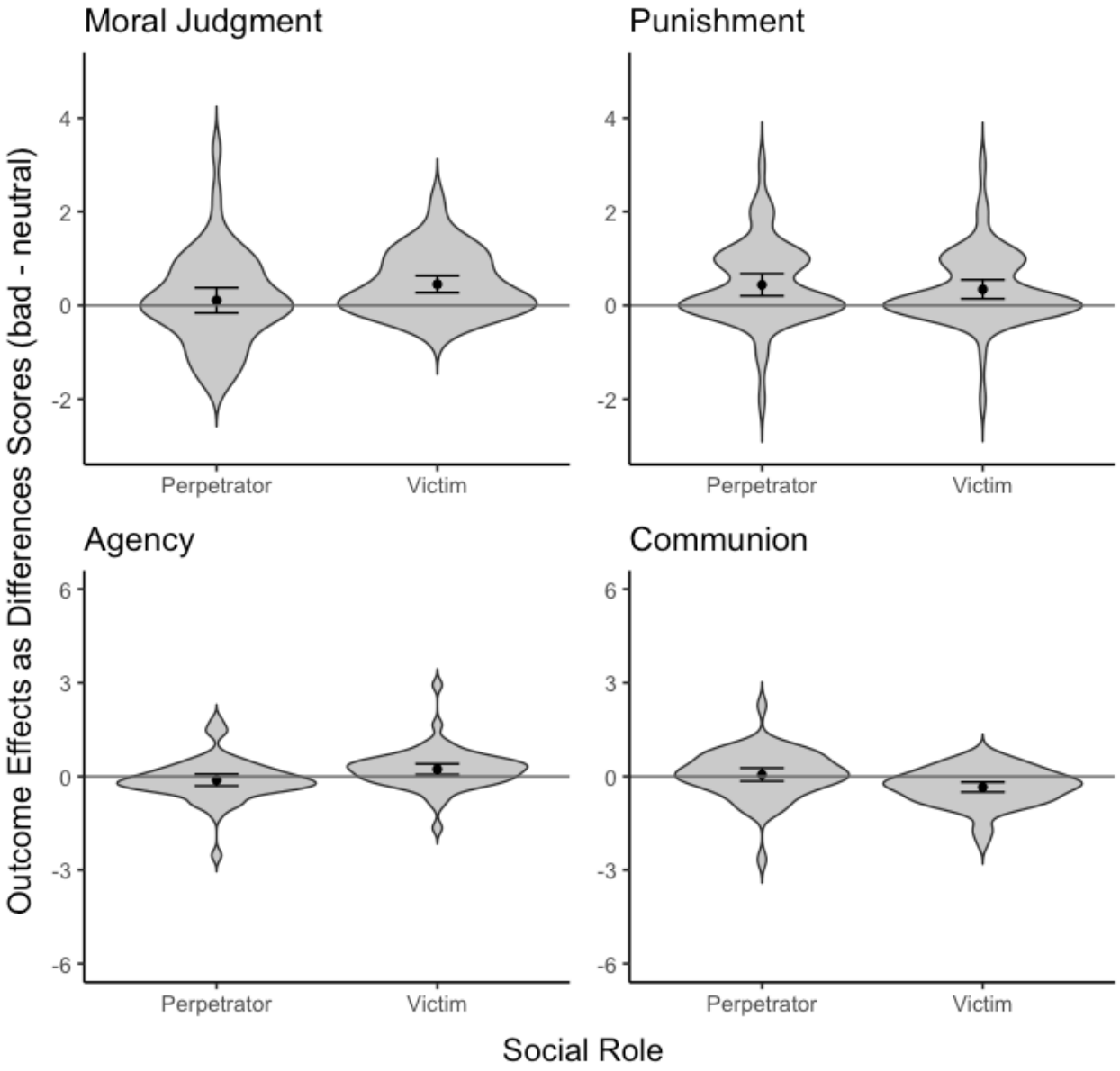

Figure 2. Violin plot of the difference scores (bad – neutral outcome) of perpetrators' and victims' moral judgments (permissibility, wrongness, blame), punishment judgments, agency, and communion (Study 1). Points represent means, error bars represent 95% confidence intervals, and the width of density plots represent the frequency of observations.

Table 1

Descriptive Statistics and Simple Main Effects of Outcome on Moral Judgment and Interpersonal Goals (Study 1)

	Perpetrator						Victim					
Variable	neutral	bad	*t*	*p*	d_z	*CI*	neutral	bad	*t*	*p*	d_z	*CI*
Moral Judgment	3.33 (1.21)	3.44 (1.14)	0.99	.326	0.09	[0.32, -0.13]	3.58 (1.05)	4.03 (0.91)	4.21	<.001	0.46	[0.64, 0.27]
Punishment	2.46 (1.21)	2.90 (1.35)	3.96	<.001	0.34	[0.53, 0.16]	2.67 (1.09)	3.02 (1.22)	3.19	.002	0.29	[0.47, 0.12]
Agency	-0.11 (0.71)	-0.23 (0.75)	-1.15	.245	-0.16	[0.09, -0.41]	0.54 (0.73)	0.77 (0.79)	2.6	.005	0.3	[0.53, 0.07]
Communion	1.47 (1.33)	1.53 (1.49)	0.38	.703	0.04	[0.18, -0.1]	0.89 (1.07)	0.54 (1.36)	-3.86	<.001	-0.25	[-0.13, -0.36]

Note. Means and standard deviations (in parentheses) are given for perpetrators' and victims' moral judgments, punishment judgments, agency, and communion after neutral and bad outcomes. Test statistics (*t*) and *p*-values, effect sizes (d_z) and their 95% confidence intervals (*CI*) are given for the simple main effects of outcome (bad – neutral).

Regarding agency, results revealed an interaction of social role and outcome ($F(1,103) = 6.93$, $p = .009$, $\eta_p^2 = .06$). As predicted, victims reported stronger agentic goals after transgressions leading to bad outcomes than neutral outcomes (see right column of Table 1). In contrast, perpetrators reported slightly weaker agentic goals after bad rather than neutral outcomes, but this difference was non-significant (see left column of Table 1). Thus, victims and perpetrators showed outcome effects in opposite directions (see lower left-hand side of Figure 2). To compare the strength of outcome effects across social roles, we tested whether the negative deviation from zero observed among perpetrators was smaller than the positive deviation from zero observed among victims. In terms of the absolute deviation from zero, we found that the difference in the strength of victims' and perpetrators' outcome effects was non-significant (b = 0.12, $t(103) = 0.98$, $p = .164$, $d = .17$).

However, analyses of communion yielded support for our hypothesis: Social role interacted with outcome ($F(1,103) = 8.73$, $p = .004$, $\eta_p^2 = .08$), such that outcome severity influenced victim's communion, being weaker when the outcome was bad rather than neutral (see right column of Table 1). As illustrated in lower right-hand side of Figure 2, perpetrators showed the reversed pattern, reporting stronger communal goals after negative rather than neutral outcomes, but again this difference was not significant (see left column of Table 1). In terms of the absolute deviation from zero, the outcome effect was stronger for victims than for perpetrators ($b = 0.31$, $t(103) = 2.41$, $p = .009$, $d = .43$). To summarize, results show the predicted outcome x social role interaction, such that victims showed larger outcome effects than perpetrators on moral judgment and communion. For agency the predicted pattern is visible though not significant.

Secondary Analyses In this section, we discuss effects of social role and mental states, which are not part of our hypotheses but interesting in their own right.

There was a main effect of social role on moral judgment ($F(1,103) = 4.34$, $p = .04$, $\eta_p^2 = .04$), such that victims judged transgressions more harshly ($M = 3.80, SD = 1$) than perpetrators ($M = 3.38, SD = 1.17$). However, for punishment judgments this main effect of social role was non-significant ($F(1,103) = 0.04$, $p = .85$, $\eta_p^2 = .00$). Consistent with the basic assumptions of the needs-based model (Shnabel & Nadler, 2015), main effects of social role on agency ($F(1,103) = 43.29$, $p < .001$, $\eta_p^2 = .3$) and communion ($F(1,103) = 9.72$, $p = .002$, $\eta_p^2 = .09$) suggest that perpetrators and victims differed in their interpersonal goals: After transgressions, victims reported stronger agentic goals ($M = 0.66, SD = 0.76$) than perpetrators ($M = -0.17, SD = 0.73$), while perpetrators reported stronger communal goals ($M = 1.50, SD = 1.40$) than victims ($M = 0.71, SD = 1.23$).

Mental states, which we manipulated as a control variable, yielded main effects on all four dependent variables ($F(1,103) > 17.16$, $p < .001$, $\eta_p^2 > .14$), such that intended actions elicited harsher moral judgments and punishment judgments, as well as stronger role-specific goals, than unintended actions. However, mental states did not interact with social role and outcome for neither of the dependent variables ($F(1,103) < 2.15$, $p > .145$, $\eta_p^2 < .02$; see online appendix for a full report of analyses). Therefore, we collapsed the two mental state conditions in the above figures and contrast analyses relating to our hypotheses.

Discussion

Study 1 provides first evidence that the outcome effect depends on the social role in which transgressions are perceived. Although many participants made identical moral judgments for neutral and bad outcomes (confirming Kneer & Machery, 2019), our results revealed a reliable average outcome effect on moral judgment among victims but not among perpetrators. By contrast, punishment judgments of victims and perpetrators were similarly influenced by outcome severity. With regards to interpersonal goals, victims indicated more agency and less communion towards the perpetrator in the bad outcome condition than when the consequence was neutral. For perpetrators, outcome effects on both agency and communion were non-significant. Overall, the results were generally in line with predictions, but not consistently supported by significant interactions, which might in part be explained by the relatively low statistical power.

Two further limitations concerning our study design restrict the conclusiveness of findings: First, we did not balance the order of the two vignettes presented to each participant. Second, because we limited the design to the social roles of perpetrator and victim, we cannot say whether the outcome effect among bystanders is weaker than among victims and stronger than among perpetrators. Study 2 was conducted to overcome these limitations by balancing the order of vignettes and including a bystander condition. Furthermore, as mental states did not interfere with the outcome effect across social roles, we omitted the mental state manipulation in Study 2. Finally, given our interest in the moderating effect of social role on the outcome effect, it was reasonable to use a scenario where outcome effects are likely to be strongest. As prior research showed that the outcome effect is most pronounced in cases of negligent behavior (e.g. Cushman, 2008), we changed the content of our vignette to a clear negligence case.

Study 2

As in Study 1, the key hypothesis underlying this extended design addressed the moderating influence of social role on the outcome effect. Being the victim in the scenario should elicit a stronger outcome effect than being a bystander. Being the perpetrator in the scenario should elicit a weaker outcome effect. Assuming that the harshness of the moral judgments determines the strength of interpersonal goals, we expected that the effects of outcome would also manifest in the parties' interpersonal goals, such that for victims and bystanders, bad outcomes elicit stronger agentic goals and weaker communal goals towards the perpetrator than neutral outcomes. For perpetrators, by contrast, bad outcomes should elicit stronger communal goals and weaker agentic goals towards the victim than neutral outcomes.

Method

The study had a 2x3x2 design with outcome (neutral vs. bad) as a within-subjects factor and social role (perpetrator vs. bystander vs. victim) and order of vignettes (neutral first vs. bad outcome first) as between-subjects factors. We induced experimental manipulations of social role and outcome in an online vignette study.

Participants We recruited a total of 300 participants on Amazon Mechanical Turk, with their IP address location restricted to the USA. After excluding participants based on our preregistered criteria (i.e. attention check, comprehension check, native language, time to complete the survey, completion of the survey; for details see online appendix), the final sample included 267 participants (141 women and 126 men, Mdn$_{age}$ = 36, range 20 – 71). This sample size affords a power of .88 to detect a medium-sized effect (η_p^2= 0.06), assuming an alpha of 1%.

Procedure After providing informed consent, participants were randomly assigned to the cells of the between-subjects portion of the design. We consecutively presented participants with two vignettes (adopted from Kneer & Machery, 2019) each of which contained a scenario in which identical negligent behavior led to either a neutral or bad outcome. For example, the vignette below represents the neutral outcome scenario we used in the bystander condition.

Beth takes care of Mary's 2-year-old son. She fills the bath, while Mary's son stands near the tub. The phone rings in the next room. Beth tells Mary's son to stand near the tub while she answers the phone. Beth believes Mary's son will stand near the tub for a few minutes and wait for her to return. Beth leaves the room for 5 minutes. When Beth returns, Mary's son is still standing near the tub. The boy then enjoys his bath.

The vignettes in the bad outcome condition included a caretaker called "Anna" (while the mother's name remained "Mary") and concluded with the sentence: *When Anna returns, Mary's son is in the tub, dead, face down in the water.* Note that names of the interaction partner only changed in the perpetrator condition, not in the bystander condition. The order of scenario presentation was randomized. In the victim condition, participants read the vignettes in the parent's role. In the perpetrator condition, participants read the vignettes in the caretaker's role. Otherwise the vignettes were identical across treatment conditions.

Measures

Moral Judgment We assessed moral judgments as in Study 1. We averaged the permissibility (reversed), wrongness and blame items after finding satisfactory internal consistency (Cronbach's $\alpha = .82$). As explained in Study 1, we analyzed punishment separately.

Interpersonal Goals We measured interpersonal goals as in Study 1. Cronbach's α were .85 for agency and .85 for communion (see online appendix for Cronbach's α of the octant scores).

Results

Preliminary Analyses To examine the effect of the order of vignettes presented to participants (neutral vs. bad outcome first), we ran 3(social role [perpetrator vs. bystander vs. victim]) x 2(outcome[neutral vs. bad]) x 2(order[neutral outcome first vs. bad outcome first]) ANOVAs with repeated measures on the second factor on the four dependent variables. Order moderated the effect of outcome on moral judgment and punishment (outcome x order interaction $F(1,261) = 4.21, p = .041, \eta_p^2 = 0.02$, for moral judgment, and ($F(1,261) = 4.11, p = .045, \eta_p^2 = 0.02$ for punishment), as well as interpersonal goals (outcome x social role x order interaction ($F(2,261) = 6.91, p < .001, \eta_p^2 = 0.05$ for agency, and $F(2,261) = 30.98, p < .001, \eta_p^2 = 0.19$ for communion). Outcome effects were considerably smaller when the bad outcome preceded the neutral outcome (for details see online appendix). As our primary interest was the moderating influence of social role on the outcome effects, we included only participants who were presented with the neutral outcome vignette first for subsequent hypotheses tests.

Hypotheses Tests[4] We ran 3(social role [perpetrator vs. bystander vs. victim]) x 2(outcome[neutral vs. bad]) ANOVAs with repeated measures on the second factor on moral judgment, punishment, agency and communion. Our key hypothesis addressed the interaction of social role and outcome. Compared to bystanders, we expected the outcome effect to be stronger for victims and weaker for perpetrators.

Starting with moral judgment, the interaction of outcome with social role was non-significant ($F(2,122) = 1.39$, $p = .254$, $\eta_p^2 = .02$). As shown in the upper left-hand side of Figure 3, and contrary to our hypothesis, outcome effects were similarly strong and significant across social roles. This was confirmed by a main effect of outcome ($F(1,122) = 70.85$, $p < .001$, $\eta_p^2 = .37$): Although, as in Study 1, the most frequent response was to make the same moral judgment across outcome conditions (41 % of participants), transgressions that ended in a bad outcome were on average judged more harshly ($M = 4.78, SD = 0.48$) than transgressions ending in a neutral outcome ($M = 4.15, SD = 0.92$).

As for moral judgment, analyses of punishment judgments revealed a non-significant outcome x social role interaction ($F(2,122) = 2.07$, $p = .13$, $\eta_p^2 = .03$), but a significant main effect of outcome ($F(1,122) = 236.82$, $p < .001$, $\eta_p^2 = .66$): Mean deserved punishment in the bad outcome condition ($M = 4.63, SD = 0.78$) exceeded punishment in the neutral outcome

[4] Following the preregistration, we tested our hypotheses with alpha = 1%, one-sided tests, without further adjustment for multiple comparisons.

condition ($M = 2.88, SD = 1.22$). The size of the effect of outcome on punishment was more pronounced than the effect of punishment on moral judgment.

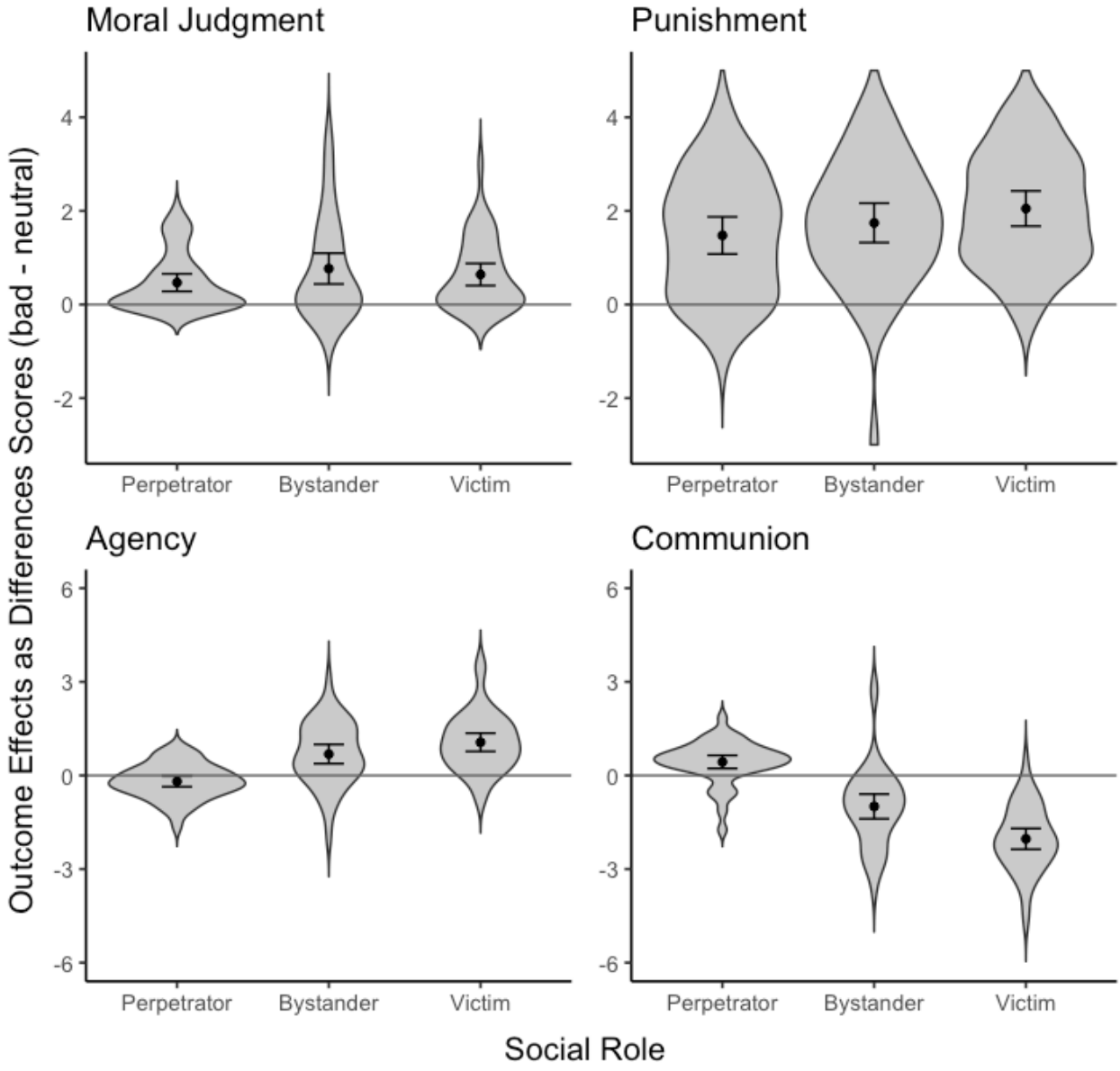

Figure 3. Violin plot of the difference scores (bad – neutral outcome) of perpetrators', bystanders', and victims' moral judgments (permissibility, wrongness, blame), punishment judgments, agency, and communion (Study 2). Points represent means, error bars represent 95% confidence intervals, and the width of density plots represent the frequency of observations.

Table 2

Descriptive Statistics and Simple Main Effects of Outcome on Moral Judgment and Interpersonal Goals (Study 2)

	Perpetrator						Bystander						Victim					
Variable	neutral	bad	*t*	*p*	d_z	*CI*	neutral	bad	*t*	*p*	d_z	*CI*	neutral	bad	*t*	*p*	d_z	*CI*
Moral Judgment	4.33 (0.72)	4.79 (0.46)	3.65	<.001	0.73	[1.06, 0.41]	3.88 (1.13)	4.64 (0.6)	6.06	<.001	0.81	[1.2, 0.42]	4.26 (0.82)	4.9 (0.22)	4.88	<.001	0.93	[1.34, 0.53]
Punishment	2.89 (1.23)	4.46 (0.81)	7.5	<.001	1.38	[1.89, 0.87]	2.70 (1.17)	4.14 (1.04)	8.97	<.001	1.64	[2.23, 1.04]	2.71 (1.20)	4.47 (0.86)	10.16	<.001	2.17	[2.89, 1.45]
Agency	-0.13 (0.57)	-0.32 (0.64)	-1.46	.147	-0.32	[-0.03, -0.6]	0.49 (0.70)	1.17 (0.99)	5.31	<.001	0.79	[1.19, 0.39]	0.66 (0.77)	1.72 (1.01)	7.96	<.001	1.16	[1.56, 0.75]
Communion	1.38 (0.78)	1.82 (1.12)	2.72	.008	.4	[0.6, 0.2]	0.81 (0.95)	-0.19 (1.23)	6.31	<.001	-0.9	[-0.48, -1.31]	1.25 (0.87)	-0.78 (1.04)	-12.46	<.001	-2.1	[-1.5, -2.71]

Note. Means and standard deviations (in parentheses) are given for perpetrators', bystanders', and victims' moral judgments, punishment judgments, agency, and communion after neutral and bad outcomes. Test statistics (*t*) and *p*-values (*p*), effect sizes (d_z) and their 95% confidence intervals (*CI*) are given for the simple main effects of outcome (bad – neutral).

Next, we tested whether social role moderated the outcome effect on interpersonal goals. Analyses yielded significant social role x outcome interactions for agentic goals ($F(2,122) = 23.91$, $p < .001$, $\eta_p^2 = .28$) and communal goals ($F(2,122) = 59.17$, $p < .001$,$\eta_p^2 = .49$). Bad outcomes elicited stronger agentic goals and weaker communal goals than neutral outcomes among victims and bystanders (see right and middle column of Table 2), whereas perpetrators reported less agency and more communion after bad rather than neutral outcomes (see left column of Table 2). As predicted, in terms of the absolute deviation from zero, we found that victims' outcome effects were stronger than those of bystanders' on both agentic ($t(122) = 2.04$, $p = .022$, $d = .39$)[5] and communal goals ($t(122) = 4.59$, $p = <.001$, $d = .89$). Also as predicted, bystanders' outcome effects were stronger than perpetrators' outcome effects on both agentic ($t(122) = 2.7$, $p = .004$, $d = .61$) and communal goals ($t(122) = 2.5$, $p = .007$, $d = .55$). The patterns of results are depicted in Figure 3 for both agency and communion.

Interestingly, the means shown in Table 2 indicate that the divergence of the interpersonal goals of perpetrators, bystanders and victims was particularly marked after transgressions that ended in a bad outcome. This result supports our assumption that victims and bystanders put greater weight on negative outcome information than perpetrators rather than responding particularly mildly after the absence of negative outcomes.

[5] According to our preregistered alpha level of 1% we are strictly speaking not able to infer differences in the outcome effect on agentic goals among victims and bystanders. The status of this hypothesis will be clarified in Study 3.

Secondary Analyses In this section we report main effects of social role on moral judgment and interpersonal goals. A main effect of social role on moral judgment ($F(2,122) = 3.85$, $p = .024$, $\eta_p^2 = .06$) revealed that victims judged transgressions more harshly ($M = 4.58, SD = 0.69$) than bystanders ($M = 4.26, SD = 0.98, t(122) = 2.45, p = .015$, $d = .38$), and so did perpetrators ($M = 4.56, SD = 0.65, t(122) = 2.99, p = .021, d = .36$). As in Study 1, the main effect of social role was non-significant for punishment judgments ($F(2,122) = 1.26$, $p = .29$, $\eta_p^2 = .02$).

Consistent with Study 1, main effects of social role on agency ($F(2,122) = 49.25$, $p < .001$, $\eta_p^2 = .45$) and communion ($F(2,122) = 32.43$, $p < .001$, $\eta_p^2 = .35$) supported the needs-based model (Shnabel & Nadler, 2015): Victims indicated stronger agentic goals ($M = 1.19, SD = 1.04$) than perpetrators ($M = -0.22, SD = 0.61, t(122) = 9.78;\ p < .001, d = 1.67$), whereas perpetrators indicated stronger communal goals ($M = 1.6, SD = 0.98$) than victims ($M = 0.23, SD = 1.40; t(122) = 7.69, p < .001, d = 1.13$). Comparisons with the bystander condition revealed that the differences between victims and perpetrators represent actual increases or decreases in agency and communion resulting from direct involvement in the moral transgression: Bystanders exhibited less agency ($M = 0.83, SD = 0.92$) than victims ($t(122) = -2.42, p = .017, d = -.37$), but more agency than perpetrators ($t(122) = 7.51, p < .001, d = 1.36$). With regard to communion, bystanders reported less communal goals ($M = 0.31, SD = 1.20$) than perpetrators ($t(122) = -7.38, p < .001, d = -1.17$). The comparison of bystanders' and victims' communal goals, however, was non-significant ($t(122) = 0.63, p = .531, d = .06$).

Discussion

Study 2 revealed that the outcome effect depends on the order in which neutral and bad outcomes are presented to participants. When the bad outcome was presented first, outcome effects on moral judgment and interpersonal goals were negligible. By contrast, when participants judged the neutral outcome first (as in Study 1), outcome effects on moral judgment emerged for all social roles. This order effect is consistent with the negative information bias, such that negative information influences subsequent evaluations more heavily than neutral information (Ito, Larsen, Smith, & Cacioppo, 1998; see also Young & Tsoi, 2013). When participants were presented with a negative outcome first, they seem to have difficulty ignoring information about the bad outcome when subsequently judging an action that ended in a neutral outcome. A fixed order of vignettes, with the neutral outcome vignette preceding the bad outcome vignette, therefore seems to be a methodological prerequisite to examine the outcome effect.

Focusing on the subset of participants who received the neutral outcome vignette first, we found that the outcome effects on moral judgment and punishment judgments were equally strong regardless of the social role participants assumed. However, consistent with our hypothesis about interpersonal goals, social role moderated outcome effects on agency and communion: Victims showed the strongest outcome effects and perpetrators showed the weakest outcome effects, with bystanders falling in between. Thus, Study 2 failed to replicate the interaction of outcome and social role on moral judgment but increased our confidence in the existence of this interaction on agency and communion.

However, the findings of Study 2 are constrained by a few methodological limitations: First of all, our bystander condition was confounded in that we asked all bystanders to imagine an

interaction with the perpetrator when indicating their interpersonal goals. Thus, it is unclear whether differences in outcome effects between bystanders and perpetrators are in fact due to stepping into different social roles as assumed by our theoretical reasoning. Alternatively, the outcome effects among bystanders might be stronger because they interacted with perpetrators and perpetrators interacted with victims. To obtain conclusive evidence, a second bystander condition is necessary in which bystanders interact with victims.

Second, the names used in our vignettes changed for the perpetrator (caretaker) among neutral and bad vignettes, where vignettes ending in a neutral outcome contained a perpetrator (caretaker) called “Beth”, and vignettes ending in a bad outcome contained a perpetrator called “Anna”. The victim’s (mother) name, “Mary”, however was constant across both vignettes. This inconsistency may have exaggerated our outcome x social role interaction: As only the victim interacted with two different perpetrators this may have enhanced their outcome effect, whereas perpetrators constantly interacted with the same victim and therefore may have shown less of an outcome effect. Study 3 was conducted to overcome these limitations

Study 3

We extended the design and hypotheses of Study 2 based on findings and methodological shortcomings of Study 2. Again, we hypothesized that social role would moderate the outcome effect. Although Study 2 yielded only partial support for this hypothesis, we upheld our theoretically justified expectation that victims would exhibit a stronger outcome effect (harsher moral judgments, stronger agentic goals and weaker communal goals when the outcome is bad rather than neutral) than bystanders interacting with the perpetrator. And accordingly, bystanders interacting with the victim should exhibit a stronger outcome effect (stronger moral judgments,

weaker agentic goals and stronger communal goals when the outcome is bad rather than neutral) compared to perpetrators.

To shed light on the mechanisms related to the outcome effect (as described in the introduction), we additionally assessed participants' estimations of the probability that the action would produce harm as well as their negligence ascriptions. By doing so we aimed to examine whether participants' perceived likelihood of a harmful outcome and their inferences to what degree perpetrators should have known their action would produce harm, mediate the effect of outcome severity on moral judgment (cf. Kneer & Machery, 2019). Although the study design does not allow for inferences about a causal relationship between negligence ascriptions, probability assessment and moral judgment as well as interpersonal goals, an inspection of the correlation between the within-subject effects of outcome serves as an indirect test of mediation.

Method

The study was a replication of Study 2 with an additional social role condition. Thus, the study had a 2x4 design with outcome (neutral vs. bad) as a within-subjects factor and social role (perpetrator vs. bystander interacting with victim vs. bystander interacting with perpetrator vs. victim) as a between-subjects factor.

Participants We recruited a total of 800 participants on Amazon Mechanical Turk, with their IP address location restricted to the USA. After excluding participants as determined in the preregistered criteria (i.e. attention checks, comprehension checks, manipulation checks, native language, time to complete the survey, completion of the survey; for details see online appendix), the final sample included 576 participants (265 women and 311 men, Mdn$_{age}$ = 35,

range 20 – 82). With this sample size, we had a power of .99 to detect a medium-sized effect (η_p^2= 0.06) using an alpha of 1 %.

Procedure

The procedure was in large parts identical with the one in Study 2 and differed only in the following respects: We held the order of the vignettes constant, always beginning with the neutral outcome. Moreover, we held the names of the protagonists constant across vignettes. And finally, we added a second bystander condition. Both bystander conditions differed solely in the assessment of interpersonal goals: Bystanders interacting with victims were asked to indicate their interpersonal goals towards the victim (Mary), whereas bystanders interacting with the perpetrator were asked to imagine an interaction with the perpetrator (Beth). Neither the vignette, nor the assessment of moral judgments differed across these two conditions.

Measures

Moral Judgments We assessed moral judgments as in Study 1 and 2. We averaged the permissibility (reversed), wrongness and blame items after determining a high degree of internal consistency (Cronbach's $\alpha = .86$). As explained in Study 1, we analyzed punishment separately.

Interpersonal goals We assessed interpersonal goals as in Study 1 and 2. Cronbachs α were .87 for agency and .93 for communion.

Probability and Negligence We assessed the perceived probability of a harmful outcome by asking participants "On a scale of 0-100 %, how high do you think was the probability of an accident? 0 % means the accident was completely improbable, 100 % means it was certain to

occur". Moreover, we assessed perceived negligence of the perpetrator's behavior by asking participants to indicate their agreement with the statement "Beth should have believed that there was a high probability of an accident" (or "I should have believed…" in the perpetrator condition) on a scale from 1 to 5.

Results[6]

Hypotheses Tests We conducted 4(social role[perpetrator vs. bystander interacting with victim vs. bystander interacting with perpetrator vs. victim]) x 2(outcome[neutral vs. bad]) ANOVAs on moral judgment, punishment, agency, and communion with repeated measures on the second factor. Our main hypothesis addressed the moderating effect of social role on the outcome effect.

Starting with moral judgments, in line with finding of Study 2 but contrary to predictions, the interaction of outcome and social role was non-significant ($F(3,572) = 0.51$, $p = .673$, $\eta_p^2 = .003$). As depicted in Figure 4, the outcome effects were similarly strong across social role conditions. As was confirmed by a significant main effect of outcome ($F(1,572) = 313.92$, $p < .001$, $\eta_p^2 = .35$), bad outcomes on average elicited harsher judgments ($M = 4.73, SD = 0.57$) than neutral outcomes ($M = 4.18, SD = 0.86$). However, in line with Study 1 and 2, the most frequent response pattern (43 %) was to make equal judgments across outcome conditions.

[6] We tested our preregistered hypotheses with alpha = 1%, one-sided tests, without adjustment for multiple comparisons.

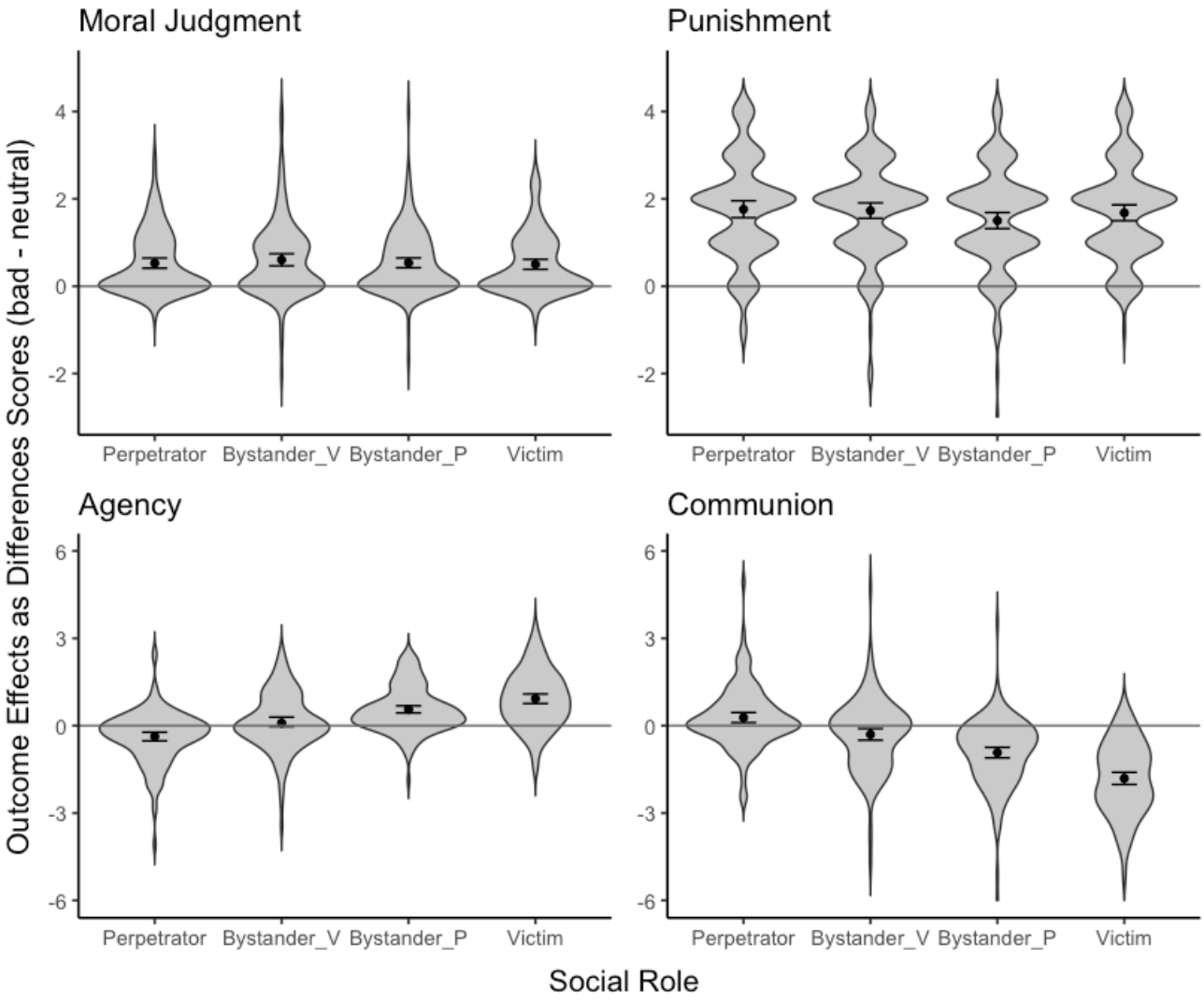

Figure 4. Violin plot of the difference scores (bad – neutral outcome) of perpetrators', bystanders', and victims' moral judgments (permissibility, wrongness, blame), punishment judgments, agency, and communion (Study 3). Bystander_V indicates bystanders interacting with the victim, whereas Bystander_P indicates bystanders interacting with the perpetrator. Points represent means, error bars represent 95% confidence intervals, and the width of density plots represent the frequency of observations.

Similarly, analyses of punishment judgments yielded a non-significant outcome x social role interaction ($F(3,572) = 1.62$, $p = .18$, $\eta_p^2 = .008$) but a significant main effect of outcome ($F(1,572) = 1278.13$, $p < .001$, $\eta_p^2 = .69$). Across social roles, punishment in the bad outcome conditions ($M = 4.46, SD = 0.85$) exceeded punishment in the neutral outcome conditions ($M = 2.80, SD = 1.08$; see Table 3a, 3b, and Figure 4). The outcome effect was more pronounced for punishment than for moral judgment.

Next, we tested whether social role moderated the outcome effect on interpersonal goals. There were significant outcome x social role interaction effects on both agentic ($F(3,572) = 50.44$, $p < .001$, $\eta_p^2 = .21$) and communal goals ($F(3,572) = 82.05$, $p < .001$, $\eta_p^2 = .30$).

As predicted, victims' increases in agency following the bad outcome were more pronounced than the increases in agency of bystanders interacting with the perpetrator ($t(572) = 3.34$, $p < .001$, $d = .42$), and the decreases in agency of perpetrators ($t(572) = 4.9$, $p < .001$, $d = .6$) confirming the result observed in Studies 1 and 2. One unexpected result apparent from the lower left plot in Figure 4 is that the outcome effect among perpetrators (decreases in agency) was stronger than the (non-significant) outcome effect among bystanders interacting with victims ($t(572) = 2.53$, $p = .012$, $d = .28$). However, as in Study 2, perpetrators' outcome effect was significantly smaller than the outcome effect of bystanders interacting with perpetrators ($t(572) = -1.71$, $p = .044$, $d = .22$).

The lower right plot in Figure 4 illustrates the pattern of outcome effects on communion across social role conditions. Statistical comparisons of simple main effects confirm the impression that the violin plot representing agency is virtually a mirror image of the violin plot

representing communion: As predicted, victim's decreases in communion were stronger than the decreases in communion of bystanders interacting with the perpetrator ($t(572) = 6.51$, $p < .001$, $d = .74$) and the increases in communion of perpetrators ($t(572) = 10.93$, $p < .001$, $d = 1.35$) confirming the findings in Studies 1 and 2. Similar to the results observed for agency, the condition of bystanders interacting with the victim did not conform to predictions, with its outcome effect being indistinguishable from the outcome effect of perpetrators ($t(572) = 0.15$, $p = .882$, $d = .02$). Also unexpectedly, bystanders interacting with the victim showed less communion following the bad outcome. However, as in Study 2, perpetrators' outcome effect was significantly smaller than the outcome effect of bystanders interacting with perpetrators ($t(572) = -4.78, p < .001, d = .59$).

In line with Study 2, these diverging outcome effects emerged primarily due to greater variation between social role conditions in the negative outcome condition (for details see Table 3a and 3b).

Table 3a

Descriptive Statistics and Simple Main Effects of Outcome on Moral Judgment and Interpersonal Goals for Perpetrators and Bystanders interacting with Victims (Study 3)

	Perpetrator						Bystander interacting with victim					
Variable	neutral	bad	*t*	*p*	d_z	*CI*	neutral	bad	*t*	*p*	d_z	*CI*
Moral Judgment	4.33 (0.77)	4.85 (0.44)	8.48	<.001	0.8	[1, 0.6]	4.06 (0.91)	4.66 (0.60)	9.93	<.001	0.76	[0.96, 0.56]
Punishment	2.88 (1.12)	4.65 (0.66)	18.57	<.001	1.88	[2.22, 1.52]	2.70 (1.07)	4.43 (0.83)	18.62	<.001	1.79	[2.08, 1.5]
Agency	0.09 (0.76)	-0.28 (0.84)	-4.69	<.001	-0.46	[-0.27, -0.66]	0.27 (0.80)	0.36 (1.14)	1.18	.235	0.09	[0.27, -0.09]
Communion	1.43 (1.02)	1.71 (1.08)	2.85	.002	0.26	[0.43, 0.1]	1.08 (1.15)	0.78 (1.53)	-3.12	.001	-0.21	[-0.07, -0.36]

Note. Means and standard deviations (in parentheses) are given for perpetrators' and bystanders' (interacting with victims) moral judgments, punishment judgments, agency, and communion after neutral and bad outcomes. Test statistics (*t*) and *p*-values (*p*), effect sizes (d_z) and their 95% confidence intervals (*CI*) are given for the simple main effects of outcome (bad – neutral).

Table 3b

Descriptive Statistics and Simple Main Effects of Outcome on Moral Judgment and Interpersonal Goals for Victims and Bystanders Interacting with Perpetrators (Study 3)

	Victim						Bystander interacting with perpetrator					
Variable	neutral	bad	*t*	*p*	d_z	*CI*	neutral	bad	*t*	*p*	d_z	*CI*
Moral Judgment	4.18 (0.93)	4.68 (0.65)	7.94	<.001	0.59	[0.73, 0.44]	4.17 (0.83)	4.71 (0.56)	9.14	<.001	0.73	[0.9, 0.55]
Punishment	2.70 (1.02)	4.39 (0.98)	17.46	<.001	1.68	[1.96, 1.4]	2.89 (1.10)	4.39 (0.90)	16.83	<.001	1.49	[1.76, 1.23]
Agency	0.73 (0.81)	1.65 (1.09)	11.5	<.001	0.95	[1.14, 0.75]	0.49 (0.88)	1.05 (1.08)	7.48	<.001	0.55	[0.69, 0.43]
Communion	1.03 (1.10)	-0.78 (1.29)	-18.14	<.001	-1.5	[-1.25, -1.75]	0.66 (1.15)	-0.26 (1.29)	-10	<.001	-0.75	[-0.59, -0.92]

Note. Means and standard deviations (in parentheses) are given for victims' and bystanders' (interacting with perpetrators) moral judgments, punishment judgments, agency, and communion after neutral and bad outcomes. Test statistics (*t*) and *p*-values (*p*), effect sizes (d_z) and their 95% confidence intervals (*CI*) are given for the simple main effects of outcome (bad – neutral).

Secondary Analyses In this section we report main effects of social role. Starting with moral judgments, we found a main effect of social role ($F(3,572) = 3.31, p = .02, \eta_p^2 = .02$), such that perpetrators judged transgressions more harshly ($M = 4.59, SD = 0.67$) than bystanders interacting with victims ($M = 4.36, SD = 0.83;\ t(572) = 3.05, p = .002, d = 0.30$) and victims ($M = 4.42, SD = 0.84;\ t(572) = 2.16, p = .031, d = .21$). The difference in the harshness of judgments between victims and bystanders interacting with perpetrators ($M = 4.44, SD = 0.75$) was non-significant ($t(572) = -0.21;\ p = .832, d = .02$). In line with the findings of Study 1 and 2, there was no significant main effect of social role on punishment ($F(3,572) = 2.26, p = .08, \eta_p^2 = .01$).

Study 3 again supported the basic predictions of the needs-based model (Shnabel & Nadler, 2015). Analyses revealed main effects of social role on agency ($F(3,572) = 65.23, p < .001, \eta_p^2 = .23$) and communion ($F(3,572) = 57.93, p < .001, \eta_p^2 = .23$). After transgressions, victims indicated stronger agentic goals ($M = 1.19, SD = 1.06$) than perpetrators ($M = -0.1, SD = 0.82, t(572) = 13.12, p < .001, d = 1.36$), whereas perpetrators reported stronger communal goals ($M = 1.57, SD = 1.06$) than victims ($M = 0.12, SD = 1.50, t(572) = 11.23, p < .001, d = 1.12$). Comparisons with the bystander condition replicated the findings of Study 2: Bystanders exhibited less agency towards the perpetrator ($M = 0.77, SD = 1.02$) than victims ($t(572) = -4.4, p < .001, d = -.40$), but more agency towards the victim ($M = 0.32, SD = 0.98$) than perpetrators ($t(572) = 4.29, p < .001, d = .45$). Bystanders also reported less communal goals towards the victim ($M = 0.93, SD = 1.36$) than perpetrators ($M = 1.57, SD = 1.06, t(572) = 5.06, p < .001, d = .52$). The comparison of victims' and bystanders' communal goals towards the perpetrator was non-significant ($M = 0.12, SD = 1.50$, for victims; $M = 0.20, SD = 1.30$, for bystanders; $t(572) = 0.6, p = .551, d = .05$).

In order to understand what could explain the impact of outcome on moral judgment and interpersonal goals, we explored the perceived probability of a harmful outcome and negligence judgments as mediators, relying on correlations of difference scores as the recommended approach for mediation in within-subjects designs (Judd, Kenny, & McClelland, 2001). As can be seen in Table 4, outcome effects on moral judgments generally correlated with outcome effects on perceived probability of harm and negligence ascriptions. The results were similar across social roles, except for the bystander interacting with victim condition, whose construct validity is, as discussed above, uncertain. Overall, our results are consistent with Kneer and Machery (2019), who hypothesize that the outcome effects on moral judgment might be driven by differences in perceived probability and negligence across outcomes. Given that the probability of the harmful outcome should be assessed *ex ante* and not *ex post,* the outcome effects on moral judgment might ultimately be a consequence of the hindsight bias (Fischhoff, 1975).

Evidence for a correlation between perceived likelihood of a negative outcome as well as negligence with punishment judgments, are weak, which is plausible due to the high sensitivity of punishment judgment to outcome information. Finally, the low and non-significant correlations with agency and communion indicate that these variables cannot be considered potential mediators of the outcome effects on agency and communion.

Table 4

Correlations of Outcome Effects on Moral Judgment, Punishment, Agency, and Communion with Outcome Effects on Probability and Negligence Judgments

	Moral Judgment		Punishment		Agency		Communion	
	Probability	Negligence	Probability	Negligence	Probability	Negligence	Probability	Negligence
Perpetrator	0.17*	0.38***	0.06	0.08	-0.05	0.15	0.07	0.05
Bystander_V	0.08	0.38***	0.08	0.35***	-0.05	0.22**	-0.01	-0.17*
Bystander_P	0.23**	0.30***	0.16*	0.09	0.05	-0.05	-0.15	-0.01
Victim	0.23**	0.35***	0.06	0.28***	0.02	-0.02	-0.06	0.04

Note. Coefficients show the correlations of the relative difference (bad – neutral) of moral judgment, agency and communion respectively with the relative difference probability and negligence judgments (Study 3). * $p < .05$, ** $p < .01$, *** $p < .001$.

Discussion

The findings of Study 3 are consistent with those of Study 2, in that the outcome effect on moral judgment and punishment judgments was equally strong across social roles. Corroborating the results of Studies 1 and 2, social role moderated the outcome effect on interpersonal goals: Victims exhibited stronger outcome effects on agency and communion than perpetrators. Furthermore, consistent with our reasoning, the outcome effect among victims was reliably stronger than the outcome effect among bystanders interacting with perpetrators. Results for the condition of bystanders interacting with the victim were not as expected. First, outcome effects were not stronger than in the perpetrator condition. Second, bystanders interacting with victims pursued less communion following the bad outcome than following the neutral outcome.

One explanation for these unexpected results might reside in the subtlety of the instructions. We first asked participants to morally judge the perpetrator's behavior (e.g. "How wrong was it for Beth to leave Mary's son alone in the above scenario"). In a second step, we asked participants to indicate their interpersonal goals towards the victim (e.g. "When I interact with Mary, it is important to me that I am friendly"). Note that the only cue for participants to understand that they should imagine an interaction with the victim was the name of the mother ("Mary"). It is possible that some participants overlooked this information, and instead imagined further interactions with the perpetrator (as they also had to assess the behavior of the perpetrator in moral regards just beforehand). This might have compromised the construct validity of this condition.

Robustness Checks Using Bayesian Regression Modeling

We ran supplementary Bayesian analyses on moral judgments as it is unclear to what extent our measures of moral judgments have the interval scale properties required for an ANOVA. In some cases, treating ordinal data as metric data can result in serious underestimation or overestimation of effects (cf. Liddell & Kruschke, 2018). Bayesian modeling allows for specifying a potentially more appropriate distribution of the outcome variable.

Using the R-Package *brms* (Bürkner, 2017), we fitted Bayesian Generalized Multilevel Models by regressing each moral judgment item (permissibility, wrongness, blame, and punishment) separately on social role and outcome, using a cumulative link function, non-informative priors, and random intercepts for participants. Across the three studies, results were almost always in accordance with ANOVA models, such that negative outcomes were judged more harshly than neutral outcomes: the Highest Density Intervals (HDI) did not include 0 (for full report of results see online appendix). With regard to the interaction between outcome and social role, in line with ANOVA models in Studies 2 and 3, these analyses generally (in 31 out of 36 analyses) confirmed the absence of evidence for a moderating effect of social role. Interestingly, Bayesian regression analyses revealed this absence of evidence for moderation also for Study 1, where the ANOVA yielded a significant interaction (see online appendix for results of separate analyses of moral judgment items). Treating moral judgments as ordinal data, as the Bayesian Generalized Multilevel Model allowed us to do, is statistically more sound than the preregistered metric ANOVAs we reported above for simplicity of presentation. Thus, these results help to resolve the inconsistent results for the social role x outcome interactions across studies by increasing our

confidence in the overall conclusion that peoples' divergent moral judgments following bad and neutral outcome are not moderated by social role.

Are Judgments and Goals Related?

Across three studies, the broad picture of results showed that the social role people assumed during a hypothetical transgression moderated outcome effects on interpersonal goals, but not the outcome effects on moral judgments. These divergent findings cast doubt on our theoretical assumption that the harshness of moral judgments would determine the intensity of subsequent role-specific interpersonal goals. Given that both moral judgments and interpersonal goals were only measured and not manipulated, it is impossible to decide whether moral judgments are causally prior to interpersonal goals or whether the reverse is true. Both causal models imply that the outcome effects (i.e., difference scores of bad outcome minus neutral outcome) on moral judgments and interpersonal goals are correlated (Judd, et al., 2001). Nevertheless, for an indirect test of within-subject mediation, we performed correlation analyses after combining the data from all three studies for greater sensitivity (for separate analyses of the three studies see online appendix). As depicted on the left-hand side of Table 5, the correlations of the outcome effects on judgments and goals were by and large small and not significantly different from zero, yielding no evidence for mediation or reverse mediation. Outcome effects on punishment and goals, however, were correlated for victims and bystanders (see right-hand side of Table 5), such that outcome-dependent changes in the willingness to punish the perpetrator were accompanied by changes in agentic and communal goals.

Looking at neutral and bad outcomes separately (see Table 6), however, revealed larger correlations between participants' moral judgments and interpersonal goals. The results

suggest that people's moral judgments positively relate to their role specific goals, across both outcomes: Perpetrators' moral judgments correlated positively with their communal goals. The more harshly they judged their own actions, the stronger the communal goals they wanted to pursue. By contrast, victims' and bystanders' moral judgment and calls for punishment were by and large positively associated with their agentic goals and negatively with their communal goals, such that the more harshly they judged a transgression, the more agentically and uncommunally they wanted to act.

Overall, then, these results suggest that (pre-existing) individual differences in moral judgment are systematically related to interpersonal goals, but that the moral judgment differences produced by the manipulation of outcome severity are unrelated to differences in interpersonal goals between neutral and bad outcomes.

Table 5

Correlations of Outcome Effects on Moral Judgment and Punishment, With Outcome Effects on Agency, and Communion aggregated across Studies 1 - 3

Variable	Correlation of Outcome Effects on Moral Judgment and		Correlation of Outcome Effects on Punishment and	
	Outcome Effects on Agency	Outcome Effects on Communion	Outcome Effects on Agency	Outcome Effects on Communion
Victim	0.1	-0.09	0.3***	-0.5***
Bystander	-0.07	-0.15*	0.15*	-0.19**
Perpetrator	-0.03	0.02	-0.04	0.05

Note. n = 233 in the perpetrator condition, n = 230 in the victim condition, n = 200 in the bystander condition (from Study 3, only the bystander interacting with perpetrator condition is reported), * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 6

Correlations of Moral Judgment and Punishment with Agency and Communion aggregated across Studies 1 - 3

	Correlation of Moral Judgment and Agency		Correlation of Moral Judgment and Communion		Correlations of Punishment and Agency		Correlations of Punishment and Communion	
	Neutral Outcome	Bad Outcome	Neutral Outcome	Bad Outcome	Neutral Outcome	Bad Outcome	Neutral Outcome	Bad Outcome
Victim	0.3***	0.41***	–0.16*	–0.49***	0.19**	0.42***	-0.34***	-0.58***
Bystander	0.29***	0.3***	–0.12	–0.18*	0.32***	0.3***	-0.24***	-0.24***
Perpetrator	0.02	–0.05	0.18**	0.17**	-0.01	-0.01	-0.08	0.05

Note. n = 233 in the perpetrator condition, n = 230 in the victim condition, n = 200 in the bystander condition (from Study 3, only the bystander interacting with perpetrator condition is reported), * $p < .05$, ** $p < .01$, *** $p < .001$.

General Discussion

The present research investigated whether people's social roles influence their moral judgments and interpersonal goals following transgressions with negative vs. neutral outcomes. We replicated the outcome effect on moral judgment among bystanders of a moral transgression (e.g. Cushman, 2008; Cushman et al., 2009; Gino et al., 2009; Lench et al., 2015; Young et al., 2010) and extended it to victims and perpetrators. Three studies yielded evidence that, on average, transgressions were judged more harshly when the scenario described a chain of events originating from the actor that due to bad luck results in harm rather than no harm to the victim. Contrary to predictions, these outcome effects were not moderated by social role. Victims and perpetrators agreed with bystanders in judging harmful transgressions more harshly than harmless transgressions. The perceived probability that the action would produce negative outcomes mediated outcome effects on moral judgment, and this indicates bias. However, we also confirmed previous reports of large proportions of participants making identical moral judgments for neutral and bad outcomes (41 %, across Studies 1 – 3; Schwitzgebel & Cushman, 2012; Kneer & Machery, 2019).[7] Viewed in

[7] According to the results of Kneer & Machery (2019), in within-subjects designs and contrastive designs, the vast majority of participants ascribe the same measure of wrongness, permissibility and blame to the lucky and the unlucky agents. The philosophical "puzzle of moral luck", they suggest, might thus be misconceived, since robust outcome effects can only be obtained in between-subjects designs (given that the puzzle draws on a *comparative* evaluation of the agents). In our experiments the proportion of participants who judge the agents identically across outcomes is considerable, though less pronounced. Importantly, however, this might be a consequence of the design choice: Rather than reading both scenarios first and reporting their judgments, participants read one scenario, judged the agent, filled in a lengthy questionnaire concerning

conjunction with small and inconsistent main effects of social role on moral judgment, these results suggest that (a) the Puzzle of Moral Luck may be confined to a barely over 50 % of participants (cf. Kneer & Machery, 2019), and (b) people in different social roles agree in their average increase in harshness following bad outcomes. This is good news for philosophers (a) and laypeople (b) alike.

Importantly, however, victims and perpetrators differed in their outcome effects with regard to interpersonal goals. These differences were strong and consistent across studies. The first difference is the nature of their responses. Consistent with the needs-based model (Shnabel & Nadler, 2015), victims had more agentic and more uncommunal goals when transgressions resulted in harm rather than no harm, whereas perpetrators had more communal and unagentic goals. The second difference is the magnitude of the outcome effect. Consistent with the prediction, derived from introspection illusion (Pronin, 2009), that victims would weight outcome information more heavily than perpetrators, the outcome effects on role-specific goals were stronger among victims than among perpetrators. Although not in every case significant, bystanders were less sensitive to outcome severity than victims, and more sensitive than perpetrators.

As might be expected based on the pattern of results reported above, the relationship between outcome effects on moral judgment and outcome effects on interpersonal goals was not straightforward. Within-subjects mediation analyses suggested that the outcome effects on moral judgment and interpersonal goals were for the most part unrelated. In other words, contrary to our initial assumption, the extent to which victims and perpetrators adjust the level

---

interpersonal goals, and were then presented with the second scenario. The contrastive aspect of the within-subjects design was thus considerably reduced.

of harshness of their moral judgment to the severity of the outcome does not predict the extent to which they adjust their goals for interacting with each other. The only exception was the relationship between victims' and bystanders' increased calls for punishment and higher agency as well as lower communion following bad outcomes.

As will be recalled, we followed Cushman (2008) by using four distinct moral dependent variables: permissibility, wrongness, blame and deserved punishment. Cushman's influential Dual Process Model separates moral judgment into two kinds. They differ in their relative weighting of information pertaining to the agent's mental states on the one hand and the causal factors (including outcome) on the other. On Cushman's view, permissibility and wrongness judgments ($Type_1$) are predominantly – though not exclusively – influenced by mental factors, whereas $Type_2$ judgments of blame and deserved punishment are strongly sensitive to causal factors such as severity of outcome. However, recent studies suggest that blame patterns with permissibility and wrongness rather than with deserved punishment, in so far as the latter is considerably more sensitive to causal factors than the other three types of moral judgment (for a discussion, see Kneer & Machery, 2019). In our experiments, too, we found the effect sizes of outcome on blame to be very similar to those for outcome on permissibility and wrongness. In Experiments 2 and 3, for instance, they ranged between η_p^2=.23 and .31 for permissibility, wrongness and blame. The effect sizes of outcome on punishment, by contrast, were about twice as pronounced (η_p^2=.66 and .69 in Experiments 2 and 3 respectively, see online appendix for detailed results).

While these results constitute further evidence in favor of the existence of two distinct types of moral judgment postulated by Cushman's *Dual Process Model*, they offer little in the way of explaining the overall mismatch between moral judgment and action disposition,

which we will label the judgment-action mismatch, or JAM for short. In the following, we would like to explore a number of possible interpretations of this result.

Experimental Shortcomings

As is well documented (Feltz et al., 2012; Kneer, 2018), it is difficult to make participants slip into the proverbial shoes of a distinct target individual so as to judge from their perspective in text-based experiments. Whereas a short set of four, somewhat abstract, questions concerning moral evaluation might not suffice to trigger role-specific judgment patterns and corresponding introspection illusion effects, things might be different as regards our measure of action dispositions. In the latter, the interactive element ("When *I* interact with *X*, it is important to *me* that […]", and hence the participant's particular role, are potentially more salient. Furthermore, the questionnaire's 32 questions circle through many possible behavioral patterns, thus exploring nuanced features of the imaginary situation. Assuming the perspective (and hence social role) of the target individual might have thus been easier for the part of the experiment that explores interpersonal goals. In short, according to our first, and perhaps least exciting, candidate explanation of the *JAM* we were only partially successful in activating perspective-taking among participants. Whereas participants did manifest the expected social role effects as measured by the CSIG rich in situational detail, the short text vignettes paired with the moral judgment questions might have failed to trigger the kinds of phenomena familiar from Pronin's research. Although this possibility cannot be dismissed completely, it is important to point out that all participants included in our analyses did answer the comprehension check regarding social role correctly.

Cognitivism

Prima facie, the statistical independence of changes in moral judgment and changes in interpersonal goals seems to spell trouble for cognitivism, according to which moral action is driven by moral judgment. But perhaps the situation is less bleak. From the point of view of cognitivism, it is a welcome finding that moral judgment is insensitive to social role: whether or not an action is wrong, and whether or not an agent is blameworthy, does not depend on whether the judging party finds themselves in the role of victim, perpetrator or bystander. All it thus takes is an explanation for the decoupling of the outcome sensitivity with regard to judgments and action dispositions. In fact, the moral luck literature provides a promising candidate for just such an explanation: agent-regret.

In his original discussion of moral luck, Williams (1981) argues that in cases where an agent is the causal origin of a bad consequence yet entirely free from moral blame, they will feel a type of regret that differs from both ordinary regret on the one hand and (morally inflected) remorse or guilt on the other. What is distinctive of this kind of situation is that, in contrast to a mere bystander, the agent "might have acted otherwise, and the focus of the regret is on that possibility, the thought being informed [in part by first-personal conceptions of how one might have acted otherwise]" (Williams, 1981, p. 27).

Agent-regret, Williams emphasizes, is not confined to the mental level but engenders distinctive behavioral dispositions: those who blamelessly harm another tend to be disposed to make amends. What is more, rather than being morally obliged, they have – as philosophers would say – special normative powers to *take* responsibility for the outcome, much like parents do for their children (Enoch, 2012). The kind of responsibility the blameless agents can decide and might be expected to shoulder, Enoch suggests, parallels the kind of

responsibility we assume in making a promise: Neither arises from a moral debt towards others, but is generated by a voluntary commitment.

If the specific social role (agent v. bystander) leaves a morality-independent normative footprint, then it will not stop short of the victim: The victim can reasonably expect the agent to manifest agent-regret, to take responsibility for the outcome and to provide restitution of some sort.

While Williams' and Enoch's reflections predict distinctive, role-specific personal goals, these are engendered directly by the causal structure, and are independent of the moral considerations invoked by the needs-based model. Agent-regret would thus provide precisely the kind of explanation needed in defense of (the possibility of) cognitivism: Moral judgments are not, and should not, be role-dependent. Action dispositions, rather than being a direct consequence of an assessment of the moral debt that the perpetrator owes to the victim (Shnabel & Nadler, 2015), might result from normative expectations regarding the voluntary commitment the perpetrator should make to redress the harm. As for victims (and bystanders), we obtained some evidence of a relationship between outcome effects and action dispositions, but this was confined to punishment judgments. What this suggests, then, is that the normative expectations regarding appropriate action of which agent regret is a symptom do not only bear on the agent of a transgression, but also to victims and bystanders (taking the perspective of the victim). Considering that we have introduced the *post hoc* explanation of normative expectations in the context of agent regret, one might think, that they apply first and foremost to agents, i.e., perpetrators. This might seem at odds with the fact that we observed the largest outcome effects not among perpetrators, but among victims. However, it is important to note that normative expectations drive the quality of the behavior (i.e., instead of the moral judgments), but not the extent to which outcome information is utilized. As we

predicted and found, victims pay more attention to outcome information when setting their interpersonal goals than bystanders or perpetrators.

Implications for Reconciliation Processes

Our results regarding the JAM have interesting implications for how the differential needs of perpetrators and victims arise according to the needs-based model. Building on the assumption that moral debt is the engine that motivates interpersonal behavior, manipulation checks used in previous research typically assessed the extent to which perpetrators acknowledged their offenses (e.g., Shnabel & Nadler, 2008) or privileged groups conceded that their privileges are illegitimate (e.g., Aydin et al., 2019). Possibly, however, the predictions of the needs-based model would also hold when agents accidentally and thus blamelessly harm others, in which case they cannot be considered perpetrators. This is an intriguing possibility which deserves further study.

Apologies are more likely to be accepted and lead to forgiveness if they are matched in length and content to the severity of the transgression (Bennet & Earwaker, 1994; Kirchhoff, Wagner, & Strack, 2012). However, the lack of connection between moral judgment and action dispositions cause concern about such a fit: Parties adapt their action dispositions in response to outcome severity differently. No matter whether actions resulted in harm or not, perpetrators always increased their communal behavior towards the victim, whereas victims only increased their agentic behavior towards the perpetrator when actions resulted in harm. Thus, our findings that victims' interpersonal goals are much more contingent on outcome severity than perpetrators' goals suggest a potential miscalibration between the type of apology expected by the victim and the one the perpetrator is offering.

Conclusion

In the present research, we observed that when victims and perpetrators reflect on a transgression, they tend to agree with each other (and with bystanders) as to their judgments of the permissibility and wrongness of the action, their (self-)blame, and the degree of punishment they call for. For some participants, these moral judgments were unaffected by the severity of the outcome, and for others, judgments were harsher when outcomes were more severe, driving the overall outcome effects that we observed. Although the unanimity of outcome effects across social roles regarding moral judgment is reassuring, our results suggest that it ends when it comes to action dispositions. Largely independent of moral judgment, victims and perpetrators were differentially sensitive to outcome information when they reported their interpersonal goals regarding agentic and communal action toward each other. Our discussion points to the fascinating possibility that moral action could be explained by role-specific non-moral normative demands familiar from the philosophical literature on agent-regret (Williams, 1981, Enoch, 2012) paired with selective information-processing predicted by the introspection illusion (Pronin, 2009).

References

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology, 93,* 751–763.

Adams, G. S., & Inesi, M. E. (2016). Impediments to forgiveness: Victim and transgressor attributions of intent and guilt. *Journal of Personality and Social Psychology*, *111*(6), 866-881.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*(4), 556-574.

Alicke, M., & Rose, D. (2010). Culpable control or moral concepts? *Behavioral and Brain Sciences*, *33*(4), 330-331.

Alfano, M., Loeb, D., & Plakias, A. (2014). Experimental Moral Philosophy. *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition).

Amaya, S. (forthcoming). Negligence: its moral significance. In Vargas, M. & Doris, J.M., *Oxford Handbook of Moral Psychology*.

American Law Institute (1985). *Model penal code: Official draft and explanatory notes: Complete text of model penal code as adopted at the 1962 Annual Meeting of The American Law Institute at Washington, D.C., May 24, 1962.* Philadelphia: American Law Institute.

Aydin, A., Ullrich, J., Siem, B., Locke, K., & Shnabel, N. (2019). The effect of social class on agency and communion: Reconciling identity-based and rank-based perspectives. *Social Psychological and Personality Science, 10*(6), 735-745.

Bakan, D. (1966). *The duality of human existence. An essay on psychology and religion.* Chicago: Rand Mcnally.

Baumeister, R. F. (1999). *Evil: Inside human violence and cruelty*. Macmillan.

Baumeister, R. F., Stillwell, A., & Wotman, S. R. (1990). Victim and perpetrator accounts of interpersonal conflict: Autobiographical narratives about anger. *Journal of Personality and Social Psychology, 59*(5), 994-1005.

Bennett, M., & Earwaker, D. (1994). Victims' responses to apologies: The effects of offender responsibility and offense severity. *The Journal of Social Psychology, 134*(4), 457-464.

Birnbaum, M. H. (1999). How to show that 9> 221: Collect judgments in a between-subjects design. *Psychological Methods, 4*(3), 243-249.

Birnbaum, M. H., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology, 45*(4), 792-804.

Bürkner, P. C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1-28.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353-380.

Cushman, F., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a “trembling hand” game. *PloS One*, *4*(8), e6699.

Enoch, D., (2012). Being responsible, taking responsibility, and penumbral agency. In Heuer, U., & Lang, G., *Luck, value, and commitment: Themes from the ethics of Bernard Williams*, 95-132.

Epley, N., & Dunning, D. (2000). Feeling "holier than thou": are self-serving assessments produced by errors in self-or social prediction?. *Journal of Personality and Social Psychology*, *79*(6), 861.

Exline, J. J., Yali, A. M., & Lobel, M. (1998). Self-serving perceptions in victim and perpetrator accounts of transgressions. Poster presented at the annual meeting of the Midwestern Psychological Association.

Feltz, A., Harris, M., and Perez, A. (2012). Perspective in intentional action attribution. *Philosophical Psychology, 25*(5): 673–687.

Falissard, B. (2012). *Psy: Various procedures used in psychometry*. Retrieved from *https://CRAN.R-project.org/package=psy*

Fincham, F. (1982). Moral judgment and the development of causal schemes. *European Journal of Social Psychology, 12*(1), *47-61.*

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*(3), 288.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from the perceived status and competition. *Journal of Personality and Social Psychology, 82,* 878–902.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review, 5,* 5–15.

Friedman, A. B. (2018). *taRifx: Collection of Utility and Convenience Functions*. Retrieved from *https://CRAN.R-project.org/package=taRifx*

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: a source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 75*(3), 617-638.

Gino, F., Moore, D. A., & Bazerman, M. H. (2009). No harm, no foul: The outcome bias in ethical judgments. *Harvard Business School NOM Working Paper*, (08-080).

Gino, F., Shu, L. L., & Bazerman, M. H. (2010). Nameless + harmless = blameless: When seemingly irrelevant factors influence judgment of (un)ethical behavior. *Organizational Behavior and Human Decision Processes, 111*(2), 93–101.

Grosse Holtforth, M., Thomas, A., & Caspar, F. (2011). Interpersonal motivation. *Handbook of Interpersonal Psychology: Theory, Research, Assessment, and Therapeutic Interventions*, 107–122.

Hartman, R. J. (2017). *In defense of moral luck: Why luck often affects praiseworthiness and blameworthiness*. Routledge.

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, *75*(4), 887.

Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the cause of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S.

Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Morristown, NJ: General Learning Press.

Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological methods, 6*(2), 115-134.

Kant, I. (1998/1785). *Groundwork of the metaphysics of morals.* (Translated by Mary Gregor.) New York: Cambridge University Press. (Originally published 1785.)

Kamtekar, R., & Nichols, S. (2019). Agent-Regret and Accidental Agency. *Midwest Studies In Philosophy*, online first.

Kirchhoff, J., Wagner, U., & Strack, M. (2012). Apologies: Words of magic? The role of verbal components, anger reduction, and offence severity. *Peace and Conflict: Journal of Peace Psychology, 18*(2), 109-130.

Kneer, M. (2018). Perspective and epistemic state ascriptions. *Review of Philosophy and Psychology*, *9*(2), 313-341.

Kneer, M., & Machery, E. (2019). No luck for moral luck. *Cognition*, *182*, 331-348.

Leary, T. (1957). *Interpersonal diagnosis of personality*. New York, NY: Ronald Press.

Lench, H. C., Domsky, D., Smallman, R., & Darbor, K. E. (2015). Beliefs in moral luck: When and why blame hinges on luck. *British Journal of Psychology*, *106*(2), 272-287.

Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. Retrieved from *https://CRAN.R-project.org/package=emmeans*

Leunissen, J. M., De Cremer, D., Reinders Folmer, C. P., & Van Dijke, M. (2013). The apology mismatch: Asymmetries between victim's need for apologies and

perpetrator's willingness to apologize. *Journal of Experimental Social Psychology, 49*, 315–324.

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong?. *Journal of Experimental Social Psychology, 79*, 328-348.

Locke, K. (2014). Circumplex scales of intergroup goals: An interpersonal circle model of goals for interactions between groups. *Personality and Social Psychology Bulletin*, *40*(4), 433–449.

Locke, K. D. (2015). Agentic and communal social motives. *Social and Personality Psychology Compass, 9*(10), 525-538.

Malle, B. F. (2005). Folk theory of mind: Conceptual foundations of human social cognition. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 225–255). New York: Oxford University Press.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, *25*(2), 147–186.

Martin, J. W., & Cushman, F. (2015). The adaptive logic of moral luck. *A Companion to Experimental Philosophy*, 190–202.

McCullough, M. E., Fincham, F. D., & Tsang, J. A. (2003). Forgiveness, forbearance, and time: the temporal unfolding of transgression-related interpersonal motivations. *Journal of Personality and Social Psychology*, *84*(3), 540-557.

Nadelhoffer, T. (2006). Bad acts, blameworthy agents, and intentional actions: Some problems for juror impartiality. *Philosophical Explorations*, *9*(2), 203-219.

Nagel, T. (1979). Moral luck. *Mortal questions*. New York: Cambridge University Press.

Nelkin, D. K. (2019). Moral Luck. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 ed.). Retrieved from *https://plato.stanford.edu/archives/sum2019/entries/moral-luck/*

Nichols, S., Timmons, M., & Lopez, T. (2014). Using experiments in ethics–ethical conservatism and the psychology of moral luck. In M. Christen, C. van Schaik, J. Fischer, M. Huppenbauer, & C. Tanner (Eds.). *Empirically informed ethics: Morality between facts and norms* (pp. 159–176). Springer International Publishing.

Noor, M., Shnabel, N., Halabi, S., & Nadler, A. (2012). When suffering begets suffering: The psychology of competitive victimhood between adversarial groups in violent conflicts. *Personality and Social Psychology Review, 16*, 351-374.

Pronin, E. (2009). The introspection illusion. *Advances in Experimental Social Psychology, 41*, 1-67.

Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology, 43*(4), 565-578.

Pronin, E., Wegner, D. M., McCarthy, K., & Rodriguez, S. (2006). Everyday magical powers: The role of apparent mental causation in the overestimation of personal influence. *Journal of Personality and Social Psychology, 91,* 218–231.

R Core Team. (2017b). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from *https://www.R-project.org/*

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review, 86,* 61–79.

Rosebury, B. (1995). Moral responsibility and moral luck. *Philosophical Review, 104*, 499–524.

Royzman, E., & Kumar, R. (2004). Is consequential luck morally inconsequential? Empirical psychology and the reassessment of moral luck. *Ratio, 17*(3), 329-344.

Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review, 22*(1), 32-70.

Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language*, *27*(2), 135-153.

Shnabel, N., & Nadler, A. (2008). A needs-based model of reconciliation: satisfying the differential emotional needs of victim and perpetrator as a key to promoting reconciliation. *Journal of Personality and Social Psychology, 94*(1), 116-132.

Shnabel, N., & Nadler, A. (2015). The role of agency and morality in reconciliation processes: The perspective of the needs-based model. *Current Directions in Psychological Science*, *24*(6), 477–483.

SimanTov-Nachlieli, I., Shnabel, N., Aydin, A.L., & Ullrich, J. (2018). Agents of prosociality: Agency affirmation promotes mutual prosocial tendencies and behavior among conflicting groups. *Political Psychology, 39*, 445-463.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). *Afex: Analysis of factorial experiments*. Retrieved from *https://CRAN.R-project.org/package=afex*

Smith, A. (1759). *The theory of moral sentiments*. London: A. Miller.

Stillwell, A. M., & Baumeister, R. F. (1997). The construction of victim and perpetrator memories: Accuracy and distortion in role-based accounts. *Personality and Social Psychology Bulletin, 23*(11), 1157-1172.

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, *59*(2), 204-217.

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal, 94*, 1395–1415.

Thomson, J. J. (1993). Morality and bad luck. In D. Statman (Ed.). *Moral luck* (pp. 195–215). SUNY Press.

Wenzel, M., Okimoto, T. G., Feather, N. T., & Platow, M. J. (2008). Retributive and restorative justice. *Law and Human Behavior*, *32*(5), 375–389.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37,* 395–412.

Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behaviour. In W. Grove & D. Ciccetti (Eds.), *Thinking clearly about psychology: Essays in honour of Paul Everett Meehl* (pp. 89–113). Minneapolis, MI: University of Minnesota Press.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wickham, H., François, R., Henry, L., & Müller, K. (2018). *dplyr: A Grammar of Data Manipulation*. Retrieved from *https://CRAN.R-project.org/package=dplyr*

Wickham, H., & Henry, L. (2019). *tidyr: Tidy Messy Data*. Retrieved from *https://CRAN.R-project.org/package=tidyr*

Williams, B. (1981). *Moral luck: Philosophical papers 1973-1980*. Cambridge University Press.

Young, L., Nichols, S., & Saxe, R. (2010). Investigating the neural and cognitive basis of moral luck: It's not what you do but what you know. *Review of Philosophy and Psychology*, *1*(3), 333–349.

Young, L., Scholz, J., & Saxe, R. (2011). Neural evidence for "intuitive prosecution": The use of mental state information for negative moral verdicts. *Social Neuroscience, 6*(3), 302-315.

Young, L., & Tsoi, L. (2013). When mental states matter, when they don't, and what that means for morality. *Social and Personality Psychology Compass*, *7*(8), 585-604.